Doug Slater
Director, Regulatory Affairs

Gas Regulatory Affairs Correspondence
Email: gas.regulatory.affairs@fortisbc.com

Electric Regulatory Affairs Correspondence
Email: electricity.regulatory.affairs@fortisbc.com

FortisBC
16705 Fraser Highway
Surrey, B.C. V4N 0E8
Tel: (778) 578-3874
Cell: (778) 214-3842
Fax: (604) 576-7074
Email: doug.slater@fortisbc.com
www.fortisbc.com

March 21, 2019

British Columbia Utilities Commission
Suite 410, 900 Howe Street
Vancouver, B.C.
V6Z 2N3

Attention: Mr. Patrick Wruck, Commission Secretary and Manager, Regulatory Support

Dear Mr. Wruck:

Re: FortisBC Inc. (FBC)

Project No. 1598987

Application for a Certificate of Public Convenience and Necessity (CPCN) for the Grand Forks Terminal Station Reliability Project (the Application)

Response to the British Columbia Utilities Commission (BCUC) Information Request (IR) No. 2

On November 19, 2018, FBC filed the Application referenced above. In accordance with BCUC Order G-43-19 setting out a further Regulatory Timetable for the review of the Application, FBC respectfully submits the attached response to BCUC IR No. 2.

If further information is required, please contact the undersigned.

Sincerely,

FORTISBC INC.

Original signed:

Doug Slater

Attachments

cc (email only): Registered Parties

Table of Contents **Page no.**

A. PROJECT NEED AND EXISTING SYSTEM

17.0 Reference: RISK OF FAILURE

Exhibit B-2, BCUC IR 2.3

GFT T1 Risk of Failure Limit

In FortisBC Inc.'s (FBC) response to British Columbia Utilities Commission (BCUC) information request (IR) 2.3, FBC states:

> FBC considers that an acceptable risk of failure (RoF) for a transmission station should be no higher than 2 percent based on industry standards. The RoF for GFT T1 is higher than this and was calculated by ABB as 2.6 percent.

17.1 Please indicate whether the 2 percent Risk of Failure (RoF) included in FBC's response applies to the entire substation or to the individual transformers.

Response:

The calculated 2.6 percent risk of failure (RoF) applies to the GFT T1 transformer. However, due to the fact that Grand Forks station has only a single 161/63 kV transformer, the RoF for the entire station is assumed to be 2.6 percent.

17.1.1 In either scenario, please discuss how this RoF compares to the industry practice.

Response:

In the discussion below, the Probability of Failure (PoF) has the same meaning as the RoF.

The CEATI report "Translating the Health Index Into Probability of Failure" states that,

> …even if it were possible to calculate a PoF accurately, an acceptable PoF would need to be determined for each individual asset. A PoF of 2% may be acceptable for a transformer located in a substation with no immediate neighbors and supplying non-critical load, but would probably not be acceptable for a transformer located in a densely populated area supplying the central business district of a major city. Other factors to consider would be the location within the

system, the redundancy in the system, the availability of a spare transformer or spare components, etc.[1]

FBC has concluded that a risk of failure higher than 2 percent for GFT station is not acceptable. ABB calculated the RoF for GFT T1 to be 2.6 percent.

Please refer to Attachment 17.1.1 for a copy of the CEATI report TI 63700-30/113.

17.1.2 If the RoF applies to the entire substation, please provide the station RoF if Oliver T1 transformer (OLI T1) was installed as a second transformer at Grand Forks Terminal (GFT) as proposed in Alternative A.

Response:

FBC has not performed a risk of failure analysis for OLI T1 and, therefore, cannot provide the RoF for the entire substation if OLI T1 were to be installed as the second transformer.

However, due to the installation of the second transformer (which meets single contingency N-1 planning criteria), the risk of a customer outage will be reduced because in the event of an individual transformer outage, the Grand Forks area load can be supplied from the second transformer.

17.1.3 If the RoF included in FBC's response applies to the individual transformers, please explain why FBC is proposing to keep GFT T1 in service despite exceeding the industry standard for RoF.

Response:

Please refer to the responses to BCUC IRs 2.17.1.1 and 2.17.1.2.

[1] CEATI International Inc., Report No. TI 63700-30/113, Translating the Health Index into Probability of Failure, page 18

18.0 Reference: OLI T1 FIELD INSPECTION ASSESSMENT REPORT
Exhibit B-2, BCUC IR 2.7, 2.9
Exhibit B-1, Section 3.4, p. 24
OLI T1 Condition and Installation

In FBC's response to BCUC IR 2.9, FBC states, "Annual lab results indicate no change in unit health and therefore FBC believes that storing OLI T1 on-site for 10 years has not negatively impacted the serviceable lifespan of OLI T1."

In FBC's response to BCUC IR 2.7, FBC states, "FBC expects the life of the two transformers to be extended if they are operated in parallel, evenly sharing the load…"

On page 24 of the FBC Certificate of Public Convenience and Necessity Application for the GFT Reliability Project application (Application), FBC outlines the following plan of transformer additions for Alternative B:

- Year 10 – Replace GFT T1 with OLI T1
- Year 25 – Replace OLI T1 with new transformer

18.1 Please confirm that FBC expects no change in transformer health to OLI T1 between now and its proposed installation at GFT in 10 years or later, if the service life of GFT T1 is extended.

Response:

Not confirmed. FBC expects some change in transformer health to OLI T1 as time progresses.

However, it is expected that with proper maintenance, the extended storage of OLI T1 at Grand Forks Terminal Station will not have a significant impact on the overall remaining unit life expectancy.

With a new transformer in place as set out for Alternative B, the risk of using the older transformers (GFT T1 and then OLI T1) is lower than using GFT T1 and OLI T1 together (as in Alternative A).

19.0 Reference: PROJECT NEED AND EXISTING SYSTEM

Exhibit B-2, BCUC IR 2.10, 2.10.1

Spare Transformers

In response to BCUC IR 2.10, FBC states “OLI T1 is currently designated as an emergency spare in the FBC system. There are two other stations in FBC’s system for which OLI T1 could potentially be used.”

In response to BCUC IR 2.10.1, FBC states:

> OLI T1 is the only designated emergency spare for the two other stations noted in the response to IR 1.2.10. FBC is currently developing a spare parts equipment strategy that evaluates the impact on system performance for the unavailability of certain major transmission equipment, including the transformers for which OLI T1 is a potential spare.

19.1 If OLI T1 is installed permanently at GFT, as in Alternative A, please explain the sparing strategy for the other two stations in FBC’s system where OLI T1 is currently designated as an emergency spare.

<u>Response:</u>

This response is being filed confidentially pursuant to section 18 of the BCUC’s Rules of Practice and Procedure regarding confidential documents adopted by Order G-15-19 because it contains sensitive system and operational information about FBC’s critical assets that, if disclosed, could jeopardize the safety, security, and operation of FBC’s transmission system.

[redacted]

[redacted]

[redacted]

[redacted]

[redacted].

[redacted]

19.1.1 Please provide the name, location, current age and risk of failure for the transformers at the other two stations where OLI T1 is designated as an emergency spare.

Response:

The transformers for which OLI T1 is designated as an emergency spare are identified in the confidential response to BCUC IR 2.19.1. As explained in the response to CEC IR 1.10.2 (Exhibit B-5), FBC does not calculate the RoF for all equipment or stations. An independent consultant will be contracted to perform the RoF calculations only if unusual trends are discovered through regular maintenance.

19.1.2 Please explain the risks to FBC's operations if OLI T1 is not available as an emergency spare.

Response:

Please refer to the confidential response to BCUC IR 2.19.1.

19.1.3 Please explain whether there are other similarly sized spare transformers in the FBC system that could be used as emergency spare transformers at GFT and the other two stations referenced in the preamble, if OLI T1 was not available.

Response:

FBC does not own any other spare transformers with a similar size and voltage level.

20.0 Reference: OLI T1 FIELD INSPECTION ASSESSMENT REPORT

Exhibit B-4, BCOAPO IR 6.1, p. 8; Exhibit B-1, Appendix D - ABB OLI T1 Field Inspection Assessment Report, p. 5

OLI T1 Condition

In FBC's response to British Columbia Old Age Pensioners' Organization et al. (BCOAPO) IR 6.1, FBC states:

> Inspection of the load tap changer by ABB revealed the possibility that acetylene was originating from the tap changer compartment. Given these findings, FBC investigated repairing the load tap changer. Based on the known history of the unit, the only realistic operation would have been an onsite load tap changer replacement. Considering OLI T1 was an emergency spare at the time, FBC deemed this approach too costly.
>
> FBC plans to replace the load tap changer if OLI T1 is refurbished and installed as the second transformer at GFT.

On page 5 of Appendix D of the Application, in the condition report for OLI T1, ABB states:

> The tap selector and contactor assembly were inspected; contact wear is normal with no sign of arcing on the main and selector contacts. Spring and contact pressure is good. Inspection of the tap changer switch components including geneva gears and drivers, push rods, bearings, levers, and operating shafts revealed no abnormal wear or defects. Inspection of mechanical fasteners revealed no loose, broken or missing components.
>
> … The motor drive mechanism appeared in generally good condition for the age of tap changer. The tap changer was operated through all positions, end stops functioned correctly, dynamic brake operated correctly, limit switches and cams are secure and operate correctly, and the drive shaft oil seal shows no signs of oil leak.

20.1 Please explain the reasons for FBC's plans to replace the load tap changer of OLI T1, with respect to the ABB inspection report or other inspection results.

Response:

The ABB OLI T1 Field Inspection Report indicated that the seal between the main tank and the Load Tap Changer (LTC) is leaking. Addressing this oil leak requires an extensive scope of work and removal of the existing LTC.

In the report, ABB does not specifically refer to the arcing contacts of the LTC. However, this component wears with each operation and should be replaced every 500,000 operations. The OLI T1 LTC has operated 653,000 times. FBC has also identified issues with the LTC motor and gear mechanism. Additionally, the OLI T1 LTC is obsolete and is not supported by its original manufacturer.

Since the LTC has to be removed to undergo exhaustive work in order to repair the issues described above, FBC considers replacing the LTC to be the preferred solution.

B. CONSULTATION PROCESS

21.0 Reference: Consultation

Exhibit B-2, BCUC IR 5.4

Indigenous Consultation

In response to BCUC IR 5.4, FBC states "FBC sent a letter on November 22, 2018 to the same list of affected Indigenous communities as included in Section 4.1.1 of the Application".

21.1 Please provide a copy of the November 22, 2018 letter sent to the affected Indigenous communities.

Response:

A copy of the letter is provided below.

Blair Weston
Community and Indigenous Relations Manager
FortisBC

FortisBC Inc.
908B Front Street
Nelson BC V1L 4C2
250-231-0176
blair.weston@fortisbc.com
www.fortisbc.com

DATE

Address

Attention: XXXXXX

As a follow up to our letter dated July 13, 2018 regarding FortisBC's proposed Grand Forks substation upgrade and line rehabilitation we would like to let you know that our Certificate of Public Convenience and Necessity (CPCN) application was submitted to the BC Utilities Commission (BCUC) on November 19th 2018.

The majority of the upgrade is installing a new transformer at the Grand Forks Terminal Station. All this work will be done within the current substation footprint. Along with the transformer replacement there will be transmission modifications in order to alleviate system constraints, maintain customer reliability, and reduce ongoing maintenance on the transmission lines.

The transmission modifications include the salvage of two power lines from Cascade Substation in Rossland to Christina Lake. The copper transmission conductor and any poles that do not have distribution underbuild can be salvaged, with the remaining structures rehabilitated. Some of the poles that will be switched to distribution are at end of life will need to be replaced which means the setting of new poles.

As a regulated utility, we must have projects like this reviewed and approved through a rigorous and transparent process with the BCUC. If FN Community wishes to register as an interested party or submit a request to intervene in the application process, information on how to get involved can be found at www.bcuc.com. All related documents filed on the public record are on the "Current Proceedings" page on the Commissions website.

If the application is approved, construction work will take place between 2020 and 2021.

If you have any questions about this project please contact me at blair.weston@fortisbc.com or at 250.231.0176

Sincerely,

Blair Weston

21.2 Please indicate if FBC has received a response from any of the Indigenous communities who received the November 22, 2018 letter.

Response:

FBC has not received any responses to its letter of November 22, 2018.

21.2.1 If responses have been received, please provide details of each response.

Response:

Please refer to the response to BCUC IR 2.21.2.

21.2.2 If FBC did not receive any response from the Indigenous communities, please indicate if FBC has followed up to ensure these communities received the notification.

Response:

FBC has not followed up on its filing notification to see if other communities that have not contacted FBC have received the notification; however, FBC is in continued discussion with the Osoyoos Indian Band (OIB) about the project. Because the OIB is the Lead Band for the Okanagan Nation, FBC determined that no follow up to the CPCN filing notification was needed.

21.2.2.1 If FBC has followed up, please indicate the method of notification (i.e. phone call, email, letter).

Response:

Please refer to the response to BCUC IR 2.21.2.2.

21.2.2.2 If FBC has not followed up, please indicate why FBC did not follow up further.

Response:

Please refer to the response to BCUC IR 2.21.2.2.

22.0 Reference: Consultation

Exhibit B-2, BCUC IR 6.1

Public Consultation

In response to BCUC IR 6.1, FBC states:

> FBC believes that a broader public consultation is not required, but rather FBC will directly contact those residents and commercial businesses that would have some limited impact during construction. As mentioned in response to BCOAPO IR 1.14.1, FBC has already begun contacting residents in the area to discuss their concerns with the project… The letters of comment in Exhibits D-1-1 and E-1 through E-6 focus on three concerns of local residents in the Copper Ridge Subdivision regarding the current and future work at the substation:
>
> a. Potential increased noise levels;
>
> b. Increase in the amount of outdoor lighting; and
>
> c. Impact on property values and resale value of properties.

FBC considers community impacts when designing and constructing substations equipment within residential areas… FBC plans to construct an engineered sound wall around the new GFT T2 transformer similar to that installed around the existing transformer to absorb and re-direct any sound away from the Copper Ridge residential area, which will minimize noise from the new transformer. Evening lighting will not increase at the substation as a result of installing the new transformer. The additional lights that will be installed during construction will only be turned on during the evening hours if an emergency occurs or crews are required to perform work during the evening hours, thereby minimizing any concerns about increased lighting.

22.1 Please indicate if noise levels will change as a result of the proposed GFT Station Reliability Project (Project).

Response:

There may be a change to noise levels at GFT on a permanent basis with the installation of the second transformer (for either Alternative A or B) as proposed in the Application. FBC is unable to quantify the potential change in noise levels at this time, as the transformer noise will vary depending on electrical loading of the transformers and operation of the cooling fans. Since FBC expects to operate both transformers (at reduced loads compared to carrying the full load on GFT T1) it is possible that transformer and cooling fan noise will be reduced. In any event, as explained in the response to BCUC IR 1.6.1 (Exhibit B-2), the additional noise levels will be mitigated to the extent possible. The transformer will include a reduced noise level specification, as is FBC's usual practice when designing and constructing substations equipment within residential areas, and FBC plans to construct an engineered sound wall around the new GFT T2 transformer similar to that installed around the existing transformer.

22.1.1 If yes, please describe in detail any change in noise levels from the current level as a result of the proposed Project.

Response:

Please refer to the response to BCUC IR 2.22.1.

22.1.1.1 Please indicate if these changes are temporary or permanent in nature.

Response:

Please refer to the response to BCUC IR 2.22.1

22.2 Please indicate if lighting at the substation will change from the current level as a result of the proposed Project.

Response:

FBC plans to add four new lights to the area around the new equipment at the station on a permanent basis. However, these lights will be used during evening construction or during an emergency event. FBC confirms that the residents will not see any increase in the lighting during normal operation.

22.2.1 If yes, please describe in detail any change in lighting that may occur.

Response:

Please refer to the response to BCUC IR 2.22.2.

22.2.1.1 Please indicate if these changes are temporary or permanent in nature.

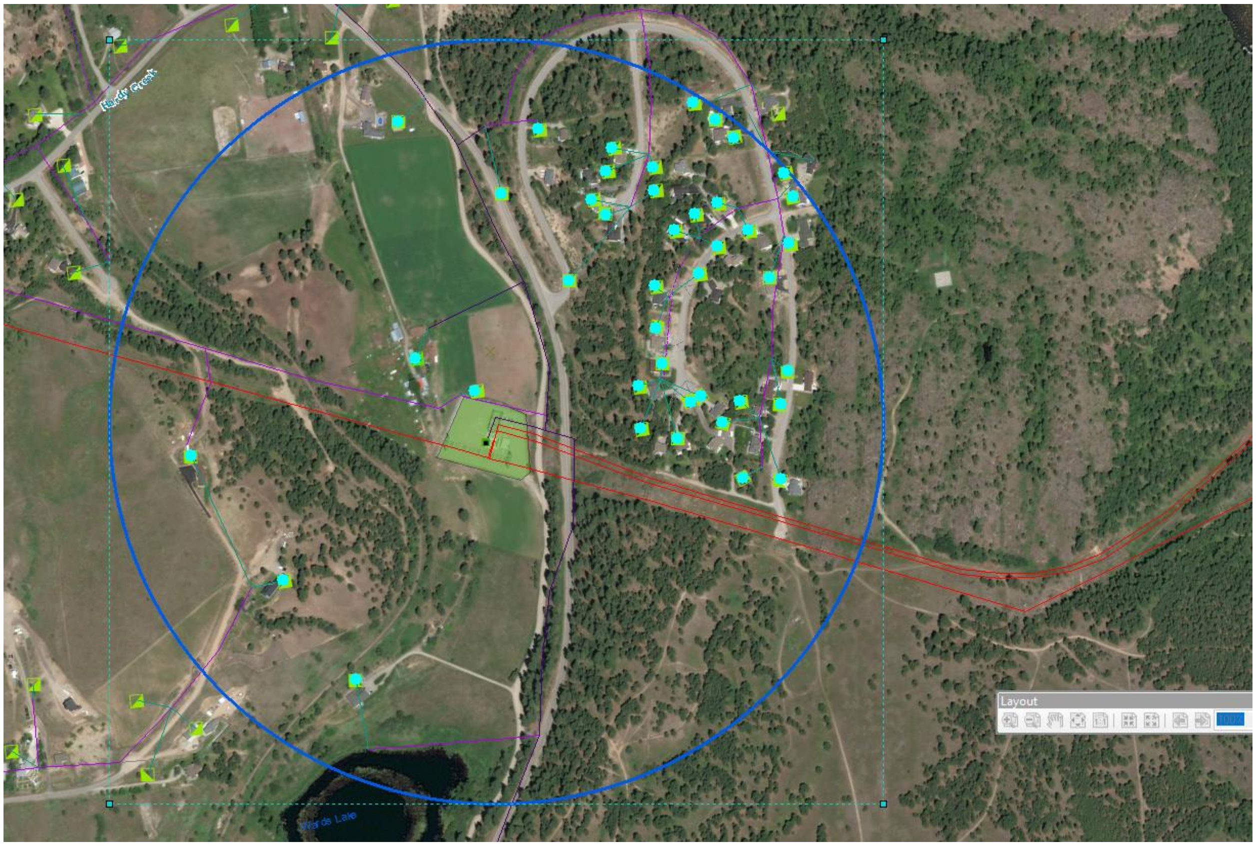
Response:

Please refer to the response to BCUC IR 2.22.2.

22.3 Please indicate how many residents and commercial businesses will be impacted by the proposed Project.

Response:

While it is difficult to determine how many residents and commercial businesses will be directly impacted by the minimal changes to the noise or lighting levels, there are 41 addresses within 250 meters of the substation. All are residential except for one belonging to the Grand Forks Irrigation District.



22.3.1 Please confirm that FBC has contacted all of the residents that will be impacted by the proposed Project to discuss their concerns.

Response:

Confirmed. FBC sent letters to all customers within 250 meters of the substation (a copy is provided in the response to BCUC IR 2.22.3.1.2). In addition, FBC contacted by telephone or voicemail those customers who filed letters of comment in this proceeding.

22.3.1.1 If not confirmed, please indicate how many residents and commercial businesses have been contacted and provide a timeline of when the other affected parties will be contacted.

Response:

Please refer to the response to BCUC IR 2.22.3.1.

22.3.1.2 Please provide a copy of communication sent to the residents and commercial businesses.

Response:

A letter sent to customers is provided below.

Blair Weston
Community and Indigenous Relations Manager
FortisBC

FortisBC Inc.
908B Front Street
Nelson BC V1L 4C2
250-231-0176
blair.weston@fortisbc.com
www.fortisbc.com

March 8, 2019

Address

Attention: XXXXXX

We are planning work in your neighborhood

FortisBC is in the planning stages of installing a new transformer at its substation on North Fork Road just outside Grand Forks BC. This transformer addition is part of a larger project needed to increase FortisBC system reliability in the Boundary region.

What does this mean for you?

During the planning process, FortisBC has heard some community concerns around perceived additional lighting and noise at the substation once the new transformer has been installed.

> **Lighting**—Lighting at the substation will not be increased. In fact, all FortisBC substations are undergoing new lighting designs, which generally decreases light pollution in the area. FortisBC will do a new lighting design as part of the Grand Forks project.
>
> **Noise**—, FortisBC plans to mitigate any new noise as much as possible. This will mean installing a transformer that is rated for noise levels within a residential area, as well as constructing a sound barrier at the station to minimize noise impacts to the community.

FortisBC is currently seeking approval from the British Columbia Utility Commission to undertake this process. If approved FBC plans to begin construction at the end of the third quarter of 2019, and is expecting to have the new transformer in service by the third quarter of 2020.

If you have any questions regarding this project, please contact me directly at 1-250-231-0176.

Sincerely,

Blair Weston

1

C. PROJECT DESCRIPTION AND PROPOSED ALTERNATIVES

23.0 Reference: ALTERNATIVES AND RECCOMENDED SOLUTION

Exhibit B-6, Industrial Customers Group IR 10.1, p. 32

Remote Disconnect Switches

In FBC's response to Industrial Customers Group (ICG) IR 10.1, FBC states:

> Adding a motor operator to the existing switches may be technically feasible. To meet the communication requirements to operate it remotely either fibre or cell communications would be required. Since there is no fibre network at these sites, it would be necessary to use cell communications. However, due to the remoteness of these areas, even cell communications may have limited reliability at these sites. Additionally, FBC has historically had issues with cell communication networks being used on remote switching applications.

23.1 Assuming FBC could meet communications requirements to install remotely-operated disconnect switches, please discuss how the installation of remotely-operated switches would affect reliability for lines 9L and 10L, and estimate any impact this would have on O&M costs.

<u>Response:</u>

Although FBC stated in its response to ICG IR 1.10.1 that the remote operation of the disconnect switches "may be technically feasible", it does not consider remote operation to be operationally preferable for the reasons described in that response. Even if remote operation were operationally preferable, FBC is unable to quantify the impact on reliability, as it would be dependent upon the causes and locations of outages. Additional O&M costs would be required due to routine and annual maintenance of such switching sites.

Attachment 17.1.1

Report for
CEATI INTERNATIONAL Inc.
1010 Sherbrooke Street West, Suite 2500
Montreal, Quebec, Canada H3A 2R7
www.ceati.com

STATION EQUIPMENT ASSET MANAGEMENT PROGRAM (SEAM)

CEATI REPORT No. T163700-30/113

TRANSLATING THE HEALTH INDEX INTO PROBABILITY OF FAILURE

Prepared by
Doble Engineering
Watertown, MA 02472, USA

Principal Investigators
Tony McGrail, Ph.D.
Ken Elkinson, P.E.,
Matthew Kennedy, M.Sc.

Sponsored by

AltaLink	**Hydro One Sault Ste. Marie**
American Electric Power	**Hydro Ottawa Limited**
American Transmission Company	**Manitoba Hydro**
Arizona Public Service Company	**National Grid**
ATCO Electric	**NRECA**
Avangrid Networks	**Newfoundland Power**
Avista Corporation	**Oakville Hydro**
BC Hydro	**Ontario Power Generation**
Chelan County PUD	**Pacific Gas & Electric Company**
Con Edison	**Powerlink Queensland**
ENMAX	**Saskatoon Light & Power**
EPCOR	**SaskPower**
Florida Power & Light	**Seattle City Light**
FortisBC	**Toronto Hydro Electric System Limited**

Technical Advisor
Ken Kopechanski

May 2018

NOTICE

This report was prepared by the CONTRACTOR and administered by CEATI International Inc. ("CEATI") for the ultimate benefit of CONSORTIUM MEMBERS (hereinafter called "SPONSORS"). Neither CEATI nor the SPONSORS necessarily agree with the opinions expressed herein.

Neither the SPONSORS, CEATI, the CONTRACTOR, nor any other person acting on their behalf (a) makes any warranty, express or implied, or assumes any legal responsibility for the accuracy of any information contained in this report or for the completeness or usefulness of any apparatus, product or process disclosed in the report, (b) accepts any liability for the use of, or damages resulting from the use of this report or any apparatus, product or process disclosed in this report or (c) represents that the use of such apparatus, product or process would not infringe upon the rights of third parties.

Furthermore, the SPONSORS, CEATI and the CONTRACTOR HEREBY DISCLAIM ANY AND ALL WARRANTIES, EXPRESSED OR IMPLIED, INCLUDING THE WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE, WHETHER ARISING BY LAW, CUSTOM, OR CONDUCT, WITH RESPECT TO ANY OF THE INFORMATION CONTAINED IN THIS REPORT. In no event shall the SPONSORS, CEATI or the CONTRACTOR be liable for any direct, special, indirect, incidental or consequential damages resulting from the use or alleged use of any information contained in this report.

ABSTRACT

This report looks at the origins and roles of Asset Health Indices (AHIs) and their use in the electric supply industry. It includes examples of systems in use, identifies the shortcomings associated with many AHIs, and suggests how to develop an AHI in a way which preserves calibrated timescales and urgency as well as allows for a derivation of an asset probability of failure (PoF).

The report also includes statistics of failures and failure modes, along with background information on the asset management framework and regulatory drivers associated with the origin and application of asset health indices.

The literature review discusses the problems generated by weighted systems not having a monotonic relationship between the score and actual condition, and thus the AHI not having a clear relationship with the PoF; the AHI must address a specific problem clearly. Furthermore, age should be an indicator of operational stress rather than a direct cause of deterioration. The use of age as a factor in condition calculations, however, is self-fulfilling: if older assets are assumed to be in worse condition and get replaced, the overall population is improved, even if replacements are made at random.

Examples of systems in practice show the influence of weighted system development, particularly using versions of a popular approach: a normalized 1–100 scoring system, where 100 is "as new". Weighted systems have a tendency to dilute any urgent condition and result in a system that is not monotonic. A worse score is not necessarily a worse asset condition, and any sense of urgency is usually lost in the calculation. The assumption that the weighted score represents an average condition does not take into account that failure does not necessarily occur on average, but rather via particular failure modes. Other systems tend to be more focused on a particular application. Many systems confuse asset condition with consequence of failure and thus become a more general indicator of a risk, as they do not supply either independent PoF or consequence of failure.

A few systems do claim to have a formulaic approach, which moves from Asset Health Index to Probability of Failure. The math and process of these systems are clear, but they are not, in the authors' opinion, valid or reasonable: this is because there are many unjustified values in the calculations; there are numerous untenable assumptions; and there are many unreferenced statements that lack authority.

An approach that uses the statistics of a population to monotonically rank assets in order of condition may be married to a known and slowly changing population to allow for categories of assets with bounds on their probability of failure, supported by history and industry statistics.

Keywords:

Health Index, Asset, Probability of Failure

ACKNOWLEDGEMENTS

This report was prepared under CEATI International Agreement No. T163700-30/113 with the sponsorship of the following participants of CEATI's Station Equipment Asset Management Interest Group (SEAM):

AltaLink	AB	Canada
American Electric Power	OH	USA
American Transmission Company	WI	USA
Arizona Public Service Company	AZ	USA
ATCO Electric	AB	Canada
Avangrid Networks	ME	USA
Avista Corporation	WA	USA
BC Hydro	BC	Canada
Chelan County PUD	WA	USA
Con Edison	NY	USA
ENMAX	AB	Canada
EPCOR	AB	Canada
Florida Power & Light	FL	USA
FortisBC	BC	Canada
Hydro One Sault Ste. Marie	ON	Canada
Hydro Ottawa Limited	ON	Canada
Manitoba Hydro	MB	Canada
National Grid	MA	USA
NRECA	VA	USA
Newfoundland Power	NL	Canada
Oakville Hydro	ON	Canada
Ontario Power Generation	ON	Canada
Pacific Gas & Electric Company	CA	USA
Powerlink Queensland		Australia
Saskatoon Light & Power	SK	Canada
SaskPower	SK	Canada
Seattle City Light	WA	USA
Toronto Hydro Electric System Limited	ON	Canada

The investigators are grateful to CEATI for the opportunity to work on this interesting issue. The constant support and guidance by the CEATI Technical Advisor Ken Kopechanski, as well as Project Monitors Dan Littlefield of Avangrid Networks, Andrew Byard of AltaLink, Jay Mearns of Pacific Gas & Electric, and Brent Jorowski and Karim Abdel-Hadi of Manitoba Hydro, was greatly appreciated by the investigators.

The authors also wish to express their gratitude for the support of colleagues, friends, and relations who have joined in conversation and debate over many aspects of this report's content.

EXECUTIVE SUMMARY

Asset Health Indices (AHIs) are becoming more common, as they can be used for asset condition ranking, allowing for short-term intervention and long term planning. AHIs use well-understood analytics on known data: the resulting AHI should therefore not be a surprise. However, an AHI is still, in fact, a model of the actual health of the asset, and to quote a statistician, "All models are wrong. Some models are useful[1]".

This report:

- Reviews the theory and practice of AHI development and deployment;
- Looks at practical examples and their shortcomings;
- Identifies issues with AHI systems in general; and
- Suggests developing an AHI that is justifiable, built on calibrated timescales for action, and which preserves a sense of urgency, thus allowing for derivation of a probability of failure (PoF).

An AHI compresses a significant amount of data into a single number in order to address a given question. In doing so, some information is lost in the hope that the resulting number provides value in and of itself. Generally, some approaches to the amalgamation of data and that subsequent production of an AHI may not be meaningful or justifiable. If an AHI is scaled 1-10, it must be clear what a score of 6 means in terms of timescale, and the data must identify a failure mode to justify an action or intervention. In addition, it must be clear how 6 relates to 7, or to other values. In many AHI systems, the scales are not monotonic in that a worse score does not necessarily reflect a worse condition or a greater sense of urgency for action. In some cases, a good score can hide a very poor-condition asset which is on the verge of failure. This report examines such systems.

There are many different ways to generate an AHI. This report reviews motivations, constraints, and the use of AHIs, noting that there are multiple possible definitions for terms such as *failure* or *health*. In addition, the variability of data is noted, as well as the paucity of good links between measurable parameters and actual failures.

A literature review shows that, not only are there many ways to develop an AHI, but also that these methods are not all similar or equally appropriate. The relationship between the input parameters and the output AHI can be tenuous, with many systems relying on an overall *average condition* approach, which uses weights to combine individual parameter or component scores. It is noted that such approaches do not retain the timescales or urgency needed for action planning and the possible derivation of a PoF. Such weighted systems can be very misleading, if not completely erroneous. Assets do not generally fail through overall condition, but through particular failure modes, and we need to track the inception and development of those failure modes if we are to be successful in planning actions. Many systems use age as an indicator of condition. This report explains, however, that the inclusion of age as a driver for asset condition must be viewed with some skepticism because replacing assets at random will usually improve the overall population condition.

---

[1] Box, G. E. P. "Science and Statistics" *Journal of the American Statistical Association*, 1976

A review of practical systems shows the breadth of approaches currently employed in trying to address an AHI development. Many practical AHI systems do not have a clear statement of intent: what, exactly, is the problem we are trying to solve? Several AHI systems seemingly obfuscate the derivation of the AHI by using complex math. In some cases, there is a sleight of hand moving from an AHI directly to a PoF—there being no clear justification of the translation from AHI to PoF. The role of AHIs in justifying expense is common, so the ability of the regulator to dissect and critique some systems is uncertain. Most AHI values the authors reviewed do not reflect the failure modes in operation or the interventions that are likely required. Furthermore, the urgency needed to address some failure modes is often hidden, lost in the averaging or dilution effects of weighted systems and possibly leading to an incorrect perception of the likelihood of failure.

This report investigates the difficulties of relating measured parameters to failure modes, and thus to AHIs and on to PoF. The field is complex and there are many opportunities that may confuse the user. Knowing the sources of variability, however, allows for the development of an approach that may have some practical justification. The created AHI will not be useful unless we retain sight of the data, deterioration, failure modes, and timescales while also retaining the sense of urgency needed for action.

The authors give one possible approach to developing an AHI that can be used to determine a justifiable and auditable PoF. The approach is somewhat heuristic, but is built on logic and can be tuned to reflect actual failure rates. The approach retains a direct link from data, through multiple analytics to failure mode identification, component analysis, and finally to an AHI which retains a sense of timescale and urgency. This allows a subsequent derivation of the PoF.

Overall, a good AHI will have an indication of timescale for action and an audit trail that allows for justification of those actions. Ideally, an AHI should be built around failure mode analysis to provide a basis for action with calibrated timescales.

An ideal AHI reflects the failure modes likely to be present and relates those failure modes to timescales; this can be developed through an analysis of known or expected failure rates. As asset populations may be small, the statistics of failures may be both imprecise and lacking accuracy. However, they at least act as a basis in fact upon which a user could develop a realistic ranking. Industry statistics, such as are available from CIGRE reports, are both useful and justifiable. It is also useful to disaggregate condition-based failures and random failures, or at least to find a proportion of the overall failure rate that is expected to be condition based.

When developing an AHI, it is important to start with a single question in mind and design an index to address that question. Start simple: not only should the index not contain surprises, it should also reflect what we already know. In addition, we can then build a comprehensible index that can become more complex with the addition of more data from a variety of sources. In developing an AHI, we are trying to model the actual health of an asset, and to quote that statistician again, "you don't get a 'correct' model by excessive elaboration."[2]

[2] Box, G. E. P., "Science and Statistics" *Journal of the American Statistical Association*, 1976

TABLE OF CONTENTS

LIST OF TABLES

LIST OF ILLUSTRATIONS

GLOSSARY

This Glossary gives a set of definitions for acronyms and terms used within the document.

Glossary of Acronyms

Item	Description
AHI	Asset Health Index
AMI	Asset Maintenance Index
APF	Asset Probability of Failure
CBHI	California Bridge Health Index
CIGRE	Council on Large Electrical Systems (Conseil International des Grands Réseaux Électriques)
CT	Current Transformer
DGA	Dissolved Gas Analysis
DNO	Distribution Network Operator
DPT	Doble PowerTest
EOL	End of Life
GSU	Generator Step Up (a transformer)
IAM	Institute of Asset Management
IEEE	Institute of Electrical and Electronics Engineers
ISO	International Standards Organization
KPL	Kenya Power and Lighting
LOAT	Life of a Transformer – a Doble Seminar
OFGEM	The UK Office of Gas and Electricity Markets
PAS 55	Publicly Available Specification 55 "Asset Management", a UK standard
PD	Partial Discharge
PMU	Phasor Management Unit
PoF	Probability of Failure
SP	Scottish Power; SP Energy Networks are DNOs
TATA	An Indian Multinational, founded by Mr. Tata
THI	Transformer Health Index
UK	United Kingdom
UKPN	UK Power Networks

Glossary of Terms

Item	Description
Heuristic	A rule based on experience
Monotonic	Uniformly increasing or decreasing

1.0 INTRODUCTION

There are many organizations using asset health indices, asset scores, and other similar methods of evaluating asset health. The aim of these indices and scores is to identify assets that are either showing signs of deterioration or are performing poorly so that intervention [1] can be applied in either the short or the long term. There are many systems currently in use that develop scores or indices and subsequently enable asset ranking. Some of these systems are based on weighted summations of component scores [1], [2]. This report discusses the advantages and disadvantages of some of these systems.

The motivation for condition assessment, asset health scoring/indexing, and subsequent intervention should first be determined. A condition assessment might be necessary to identify the assets that are most in need of replacement, or to identify assets that could benefit most from major refurbishment. In each case, the approach and the result should be different. A high-level overview of all assets may enable a more focused analysis of suspect assets.

Defining failure can cause confusion. Is failure related to system interruption [3], [4], or to a need for removal from the present asset location [5], [6]? Probability of failure (PoF) can only be successfully addressed once there is a consistent and well-understood definition. Different organizations may approach the definition of failure in different ways, according to their business needs.

Similarly, defining health (or condition, assuming that the two are synonymous) is not easy. Does asset health refer to:

- Fitness for function?
- Level of deterioration of one or more components?
- Likelihood of unplanned unavailability?
- Likelihood of needing replacement?
- Probability of failure?
- Something else?

If an asset manager is hoping to assess the PoF, then the definition of health used for the health assessment must relate to the agreed-upon definition of failure.

It is also necessary to consider the accuracy of a health assessment or review. The data used for assessments, which often represent average condition, has an associated level of uncertainty, which is often high. Multiple elements have inherent variability, such as the use of judgment, data interpretation, etc. The concept of an average condition must also be reviewed, as assets tend to fail through specific failure modes rather than through overall deterioration.

Relating observations or measured parameters to the actual condition/degradation of the asset is not simple, easy, direct, or obvious. Assessments normally compare the results or observations to acceptable levels. As assets are not typically run to failure, especially when they are continually monitored, there is a lack of adequate failure statistics that can be related to assessment data, meaning that it is difficult to accurately assess when failure is likely to occur. Furthermore, a result of a test that might indicate the imminent failure of one asset may be an acceptable result for a similar asset (e.g., one from a different manufacturer).

Condition assessments become more meaningful when the gap between what is predicted and what is actually found is closed. This requires the dismantling of units that have been removed from service to confirm their predicted condition. There are cases [7] where forensic analysis shows some positive correspondence between predicted and actual conditions, but also cases which show some variability.

Furthermore, the PoF is often related to system and external conditions. The PoF of both a circuit breaker and a transformer would be a function of the frequency of faults and the fault level at the substation.

Even if it were possible to calculate a PoF accurately, an acceptable PoF would need to be determined for each individual asset. A PoF of 2% may be acceptable for a transformer located in a substation with no immediate neighbors and supplying non-critical load, but would probably not be acceptable for a transformer located in a densely populated area supplying the central business district of a major city. Other factors to consider would be the location within the system, the redundancy in the system, the availability of a spare transformer or spare components, etc.

1.1 Overview: Motivations for Condition Assessment

CIGRE Brochure 660, from Working Group B3.32, is nominally concerned with optimized maintenance of air-insulated substations [6]; however, it gets to the heart of the need for maintenance and the need to understand asset condition. The text notes that many organizations have "aged network assets that require increasingly more maintenance and capital expenditure to sustain levels of service."

The brochure notes the need for asset management, which includes a greater emphasis on condition and diagnostics, not just for maintenance but also to provide optimal solutions for "individual business needs and drivers".

Such condition and diagnostic information applies equally to maintenance decisions, loading contingencies, spares policies, and so on. Thus we have arrived at the heart of Asset Management, as described in Asset Management standards PAS 55 [8] and ISO 55000 [9], but neatly encapsulated in Brochure 660: "The problem is, therefore, how to manage sustainment of the network infrastructure to ensure a smooth passage into the future and to do this at optimized cost, risk and performance" [6].

Condition assessment and ranking supports the business of the organization, as discussed in "Asset Management: an Anatomy" [10]. Translating a condition assessment into a probability of failure (PoF) supports quantitative and detailed risk calculations, assuming that we can achieve such a translation effectively.

Some systems develop health scores out of broadly defined principles, and some systems employ extremely detailed and prescriptive criteria (see Section 2.0 for a literature review on this topic and Section 3.0 for examples of these systems). The business value of a score/index to an organization depends on its use, justification, and auditability. A new draft guide for DNOs (Distribution Network Operators) in the UK, for example, is comprehensive and prescriptive, but may not lead to a valid PoF (see section 4.16).

1.2 Asset Probability of Failure

1.2.1 Definition of Failure

Defining failure is not necessarily a simple matter, but it is essential if a user wants to determine the PoF. A first attempt at defining a failure might be to suggest that, if an asset causes a system interruption, then it has failed. An alternate definition might relate failure to a condition assessment indicating that the asset is not fit for its purpose. Table 1-1 summarizes situations where an asset causes an interruption or is determined to not fit its purpose.

Table 1-1: Asset Trips and System Interruptions

Asset Trips or Causes System Interruption	Asset Does not Trip or Cause System Interruption
Problem with Asset – Cannot be repaired for return to service	Assessment indicates asset is not fit for purpose – Cannot be repaired for return to service
Problem with Asset – Can be returned to service after major repair	Assessment indicates asset is not fit for purpose – Can be returned to service after major repair
Problem with Asset – Can be returned to service after minor repair	Assessment indicates asset is not fit for purpose – Can be returned to service after minor repair
External Cause (e.g. animal on bushings) – Results in consequential damage – Cannot be repaired for return to service	Assessment indicates asset is not fit for purpose – Fit for purpose in another location (e.g. substation with lower loads or fault level)
External Cause (e.g. animal on bushings) – Results in consequential damage – Can be returned to service after major repair	Assessment indicates asset is not fit for purpose – Fit for purpose in another location – but not economical to relocate
External Cause (e.g. animal on bushings) – Results in consequential damage – Can be returned to service after minor repair	
External Cause (e.g. animal on bushings) – No consequential damage	

When an asset causes an interruption to the network and cannot be repaired and returned to service, then a failure has occurred. Most analysts would also likely agree that an interruption resulting from an external cause that does not result in consequential damage would not be considered a failure. However, several of the other cases shown in the table are not as obvious and could lead to an extended debate.

Further questions arise from consideration of the table. If an animal on the bushings caused damage to one or more of the bushings, is this a failure of the asset or the bushings? Some organizations take bushing failures to be separate from transformer failures. If spare bushings were not available and the asset was therefore scrapped, does this change the answer? As a loose analogy, the tires on a car may need replacement or attention, but that does not require the replacement of the whole car.

A problem (determined by interruption or assessment) that requires major repair would often result in repair in the case of a newer asset, but in scrapping in the case of an older asset since the repair would be considered uneconomical. Many would consider that an asset had failed if it were scrapped, but may not consider the newer asset (with possibly the same problem) to have failed because it was repaired. The economic decision is often a function of the regulatory regime in operation and the Organization's ability to manage Capital and Operational expenses, but the decision on financial rather than technical grounds may ultimately affect failure statistics.

In some organizations, the requirement to remove a transformer from its stand to repair, refurbish, or replace it is interpreted as an indicator that a failure has occurred. If the same repair or refurbishment work occurs in situ, a failure has not occurred. However, the regulator for the industry may define failure as a need to spend Capital money[3], regardless of the location of the asset where the capital is spent.

Some organizations also include *soft failures* in their failure statistics. This refers to situations where assets are assessed as being likely to fail, removed from service, and replaced before a hard failure occurs. Both hard and soft failures are condition-based, but one occurs in real-time while the other is anticipatory.

ISO 18095 [11] applies to power transformers and provides valuable guidelines for the use of terms. It is not yet widely used, but as it is part of the ISO suite of documents it is likely to grow in popularity and influence.

1.2.2 Asset Failure Statistics

Once failure is defined, data can be collected which will allow for the development of failure statistics; however, a standardized approach to data collection is not yet used across utilities [4].

The identification of the asset that failed and the cause of failure can be difficult and requires forensic analysis and engineering judgement [3]. Many utilities are reluctant to spend the money to perform a detailed forensic analysis of a failed asset. An older asset is often expected to fail, and since the asset has little or no residual value, it is difficult to justify the expense of a detailed forensic analysis when failure does occur.

In these cases, the cause of failure may be little more than a guess, often made by an engineer who has limited knowledge of the asset. The assumed failure mode is often based on available condition assessment information: for example, if a transformer with a high level of furans failed, it may be assumed that the cause of failure was deterioration of the cellulose. This possibly incorrect assumption becomes part of the failure statistics and justifies the relationship between the condition assessment information and PoF due to the assumed failure mode.

CIGRE Technical Brochure 642 surveyed 56 utilities in 21 countries, covering 964 major transformer failures between 1996 and 2010, in a pool of 167,459 transformer years. Note that a definition of failure was provided at the time of survey and is detailed in the brochure. It is essential that users of the information provided in the brochure understand the definition that has been used for the collecting of this statistical information (see Appendix A).

[3] Taken from private correspondence discussing a utility's policies and regulatory environment.

The failed transformers were manufactured from the 1950s to 2009 [3]. The failure rates by voltage are described in the following table taken from the Technical Brochure. The survey showed the failure rates had less than 1% variation between different applications (i.e., generation, transmission, and distribution).

Table 1-2: Indicative Asset Failure Rates (CIGRE Technical Brochure 642)

INVESTIGATED POPULATION AND FAILURE RATES OF SUBSTATION TRANSFORMERS

POPULATION INFORMATION	HIGHEST SYSTEM VOLTAGE [kV]						
	69 ≤ kV < 100	100 ≤ kV < 200	200 ≤ kV < 300	300 ≤ kV < 500	500 ≤ kV < 700	kV ≥ 700	All
Number of Utilities	11	38	31	27	3	4	58
Number of Transformers	2,962	10,932	4,272	3,233	434	348	22,181
Transformer-Years	15,267	64,718	37,017	25,305	4,774	2,991	150,072
Major Failures	144	280	186	152	27	10	799
FAILURE RATE	0.94%	0.43%	0.50%	0.60%	0.57%	0.33%	0.53%

The report notes that "All populations show a low hazard rate and no distinct bathtub curve character" and that, in the population surveyed, age related deterioration could not be identified; unusual system events are suggested as a more likely cause[4]. The brochure also notes that the replacement of older units that have not failed biases population statistics. The overall lack of age-related issues implies that purely time-based interventions will be less beneficial, and that asset condition information is required to focus resources.

Two of the contributors to the Brochure also note the importance of the process for statistical data acquisition and the importance of proper data collection and analysis. The Brochure also quotes data from a particular utility [4]: this data is for failures recorded between 1996 and 2006, and indicates a rise in failure rates with age, as shown in Figure 1-1. Whether or not this trend applies more broadly will depend on whether other organizations operate and manage their assets in the same way as the studied utility.

[4] In Nowlan and Heaps' original work on RCM, they identify six failure profiles, the bathtub curve being one of them. They also note that the bathtub curve only applied to about 4% of the assets they reviewed [55].

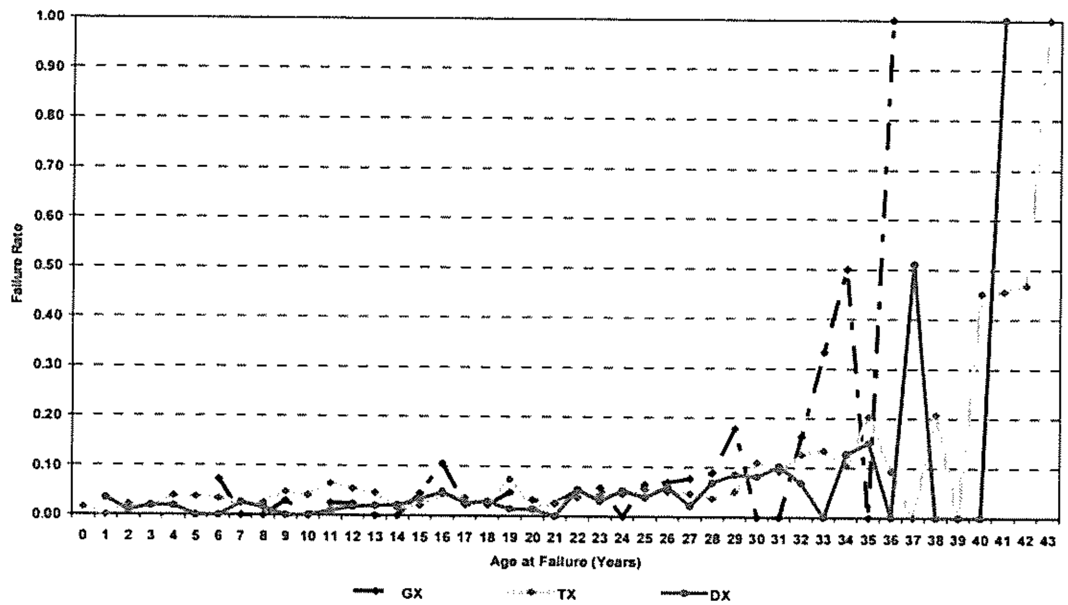


Figure 1-1: Hazard Function for Generator Step Up, Transmission and Distribution Transformers

The most common locations for failures depended on the application: GSUs having insulation issues over 30% of the time, transmission units having protection issues over 25%, and distribution units having over 35% for windings and >35% for unclassified issues. Organizations should consider the differences between the reported data and data from other sources, and question the relevance of any of the statistics to their own situation.

1.3 Failure Modes

Failure modes may be broadly classified into two types: internally caused and externally caused [5]. A poor connection inside the asset leading to looseness, arcing, and subsequent flashover would be an internally caused failure. A failure caused by multiple close-in lightning strikes leading to flashover would be an externally caused failure. The two failures, however, are related. As insulation deteriorates with time and operation, an external event that for a relatively new asset would have caused no major issues may now lead to failure, as shown in Figure 1-2 from CIGRE Technical Brochure 227, "Life Management Techniques for Power Transformers." This brochure is further discussed in Appendix B.

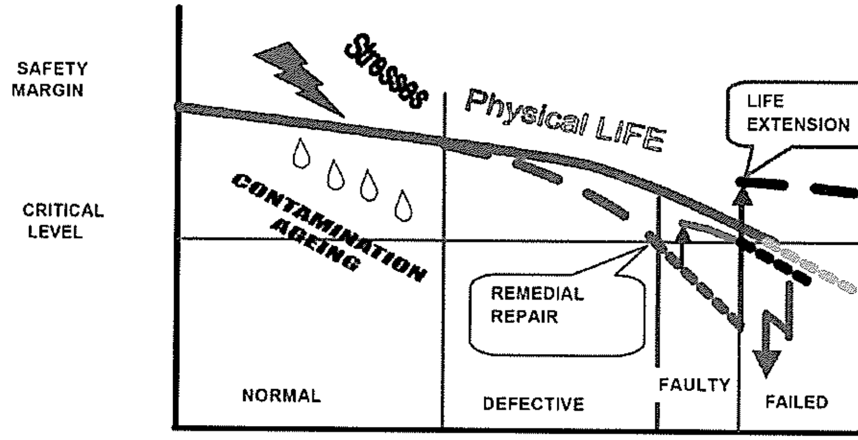


Figure 1-2: Asset Deterioration
From CIGRE Technical Brochure 227

As previously indicated, there are cases [7] where the forensic analysis of assets that have been removed from service due to poor condition has shown some correspondence between predicted and actual condition, but has also shown some significant variability. The forensic analysis of assets that are retired due to soft failures allows for a better understanding of how known condition assessment techniques relate to failure modes of the asset; however, if soft failures are included in failure statistics and a detailed forensic analysis of the soft failures has not occurred, incorrect statistical evidence may be produced, ultimately creating a false sense of accuracy in assessing failure modes.

1.4 Expected Life - Analytics

What is the expected life of a transformer, or of a circuit breaker?

If the expected life of a transformer or circuit breaker cannot be specified on the day of installation, it will be difficult to calculate the remaining life later. Some organizations have tried to specify the expected asset life and the expected age at which deterioration leads to poor performance, but these attempts have had mixed success.

An asset condition assessment assists in identifying the possible condition, while an asset health index allows the assets to be ranked to highlight those in the worst condition. As indicated in [12], however, the asset health index is an estimate at best, and therefore any analysis of remaining life is also an estimate.

1.5 Business Context and Drivers

The aim of an asset health score/index is to improve decision-making and manage risk, but the timescales for the application of a score/index may vary. The individual organization must clarify and define the nature of the score/index [1], [3], [13].

1.6 Standards and Related Documents

There are several documents and standards relating to Asset Management that are relevant to the discussion [8], [9], [11]. These documents provide a discussion of Asset Management and its need to balance risk, cost, and performance. In terms of operational strategies, an Institute of Asset Management document covers many of the elements [14]. These documents provide a useful framework for asset management review.

ISO 18095 [11] has already been noted in this report: it digs deeply into failure modes, measurable parameters, and ways to identify deterioration and degradation within a power transformer. The document does not deal specifically with health indexing but does discuss how to choose the best possible attributes, test results, and condition monitoring combination to prevent a power transformer failure. This standard is an excellent model for a discussion of failure modes and symptoms in power transformers. The final step of health indexing is not included in this document, but the foundation of knowledge that is presented is integral to AHI design and not discussed to this level of detail in other standards or articles.

"Asset Management: an Anatomy" [10] is a useful document when implementing Asset Management systems. In particular, section 6.6.5 covers "Asset Performance and Health Monitoring," and has useful commentary on both asset health and the involvement of stakeholders:

> The term "asset health" is used in relation to measures that monitor the current (or predicted) condition or capability of an asset to perform its desired function, by considering potential failure modes…It is important to review the cost effectiveness of monitoring – and involve both operations and maintenance personnel as many failure modes can be detected by operators. [10]

The Institute of Asset Management has a suite of documents relating to the implementation of Asset Management and Subject Specific Guidance documents, each written by practitioners of Asset Management from a variety of industries [14].

The subject-specific guidance includes a discussion of data and analytics and the method needed to improve results. It notes that improving data quality is an admirable task, but experience has shown that the most effective approach is to "analyze and understand the information you have" [14]. The route to good decision-making and problem-solving passes through a good analysis of poor data, as shown in Figure 1-3. Good analysis of poor data may be adequate for many problem solutions.

		Data Quality	
		Poor	Good
Analytical Ability	Good	Route for success	Target Capability
	Poor	Initial status	Avoid this route

Figure 1-3: The Route to Good Analytical Capability

1.7 Getting Started

Before starting the process of condition assessment or the development of an AHI, the first question to answer should be: "What problem are you trying to solve?" [1]. If this can be answered clearly and precisely, a measure, score, or index can be set up which helps solve the problem. The necessary data sources, assumptions that need to be made, the analyses that need to be applied, and the meaning of the results that will be obtained can also be determined.

Without a clear question, any answer will be uncertain. Even with a clear question, the answer may be based on poor data, poor analyses, poor assumptions, etc. The answer (i.e., the AHI) may be convincing, but may be equally unclear.

If an AHI is developed to identify the assets that most need refurbishment (e.g., repair of minor oil leaks, repair of rust), the business drivers related to the refurbishment of assets need to be identified. For example, if a utility had a policy that precluded the refurbishment of assets over 40 years old, the asset age could be included in the AHI. Similarly, if an AHI value was required to identify the assets that most needed replacement, the business drivers related to replacement would need to be identified.

Not all business drivers need to be included in the AHI, however, and it may often be more appropriate for some business drivers to be treated separately. As noted in Section 1.0, many factors will probably be considered before deciding to replace an asset. For example: the location; the criticality of the load; the availability of spare parts; and possibly the depreciated value of the asset, which may be a major factor in a regulated environment. It is possible to develop an index that includes all of the business drivers; however, this index would no longer be related only to the health of the asset, but also to the operational value or criticality of the asset, and thus could not be clearly related to PoF. An alternate approach would be to develop an AHI that only related to those aspects of asset health that drives replacement, and deal with the other business drivers separately, possibly with another index. There is no rule, as far the authors are aware, which precludes having multiple indices for multiple applications.

It is essential to understand that not all aspects of health directly relate to replacement. Health issues that can easily be addressed by maintenance would not generally be drivers for replacement, and would therefore not be included in a replacement index. Similarly, health issues that cannot be corrected through maintenance are generally not included in an index designed to identify assets that require corrective maintenance. Therefore, these indices, if correctly designed, cannot be directly related to PoF.

In practice, many users develop an AHI for fleet screening to allow them to identify the assets that require further investigation or assessment. An AHI for fleet screening is generally developed using data and information that is readily available for most assets of that type.

For example, some users may develop an AHI for transformers based only on the results obtained from testing oil samples. The small percentage of transformers that rate poorly in the AHI may then be scrutinized in more detail using in-service test tools, such as PD or IR, or may be taken off line for further inspection and testing as well as a full condition assessment.

Although oil analysis and dissolved gas analysis (DGA) of an oil sample from the main tank might be very good indicators of many common failure modes, they will not usually provide any indication of problems with the bushings or the tap changer; however, a transformer AHI developed for fleet screening could also include bushing test results, tap changer maintenance records, and information from site inspections, as well as details of age, loading, family history, etc. The question that is asked guides the data needed to develop an index.

AHIs based on minimal information might still be useful asset management tools, but it is unrealistic to expect that the results of the simple AHI could be related in detail to the PoF.

Some practical systems outlined in this report have very clear questions and responses; however, the responses may be very inaccurate and imprecise. It appears that in some cases, the process of producing an AHI for business purposes has become more important than the details of the process itself. Asset managers may have lost sight of the original aims of a health index by developing a system that produces a result but does not clearly address the question posed.

For an AHI to be closely linked to a PoF, it should be designed with that goal in mind, built on the failure modes that relate to the definition of failure, and should use measurable parameters and good statistics.

Statistics are necessary because all natural phenomena are described by appropriate statistics. In order to measure the breakdown voltage of a particular air gap, there are many controls that must be put in place, including voltage source, accuracy of measurements, shape and cleanliness of gap forming materials, and contents of the air (from molecular content to particle count). The result is a likely voltage, but also a range [15]. The actual breakdown voltage will be described by a statistical distribution, with a confidence level associated with the result lying in a particular range. Statistical distributions cannot be escaped.

1.8 Asset Health and Asset Probability of Failure

When generating an Asset Health Index or Score, the actual condition of the asset is estimated because the data and analytics cannot be perfect or perfectly up-to-date. Furthermore, some failure modes cannot be detected through testing, observation, or analysis of other forms of data.

1.8.1 Asset Components

These are some of the aspects to consider:

- An asset may have one or more components.
- Each component may have several failure modes.
- Failure modes may apply to several components.
- Most failure modes have symptoms; symptoms may apply to several failure modes, so identifying the failure mode may be difficult.
- Some failure modes do not have obvious or measurable symptoms until a failure has occurred or is very near to occurring.
- Measurements may be made on components; but some measurements include a number of components.
- Measurements may indicate the presence of a symptom.

- Most measurements only allow us to assess the average condition of the component being measured.
- Other information is useful in assessing asset condition. Other information may include knowledge of the design, understanding of previous problems, and knowledge of how the asset has been used, such as having been subjected to heavy loading, over-voltages, or fault currents.

A symptom is an indication of a failure mode. A number of measurements or observations may be required to confirm a symptom, and the presence (and absence) of a number of symptoms may be used to determine which failure mode is most likely occurring.

Human illness can be a helpful analogy for understanding how symptoms relate to failure modes. In a human, elevated body temperature is a symptom of many illnesses. When a doctor observes other flu-like symptoms in conjunction with a high temperature, the best diagnosis would probably be that the patient has the flu; however, the symptoms may also indicate other ailments. Note that the doctor would probably also consider the patient's medical history, family history, and other information when diagnosing the problem. If the patient had just returned from overseas, had recently been sick, or was genetically predisposed to a disease, the doctor may consider another diagnosis and possibly arrange for tests and follow-up assessments. Note also that the absence of some symptoms is often critical to the diagnosis. The doctor would probably need to ascertain that some of the symptoms of pneumonia were absent before determining that the patient had the flu.

The assessment of an asset would follow a similar process. The relationships between symptoms and assets are summarized in Figure 1-4, using one-to-many and many-to-many relationship indications to illustrate the complexity of the situation.

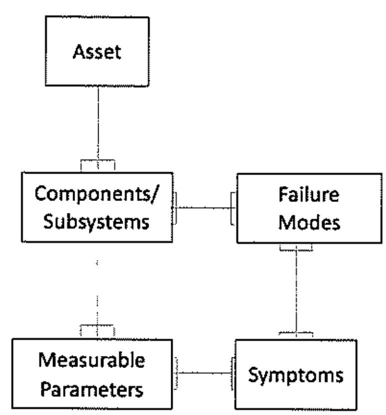


Figure 1-4: Asset and Symptoms Relationships

Figure 1-4 illustrates that an asset may have one or more components (or sub-systems), each component may have one or more failure modes, and each failure mode relates to one or more symptoms; however, the symptoms may also relate to one or more failure modes. For a small number of entities, the relations are quite complex. If this data was to be recorded in a relational database, the many-to-many relationships, such as failure modes to symptoms, would be broken out in a link table to manage the data redundancy.

A medical analogy can again be used to illustrate this point:

- A human has an appendix as a component.
- The appendix may become enlarged and infected, leading to peritonitis and, ultimately, death of the human.
- One symptom may be fever.
- A measurable parameter could be temperature.

The many-to-many relationship illustrates that fever is not a unique symptom of appendicitis, and appendicitis does not have fever as its only symptom.

When applied to a transformer:

- A transformer may have a bushing as a component.
- The bushing may have a failure mode based on stress grading foil erosion.
- The symptom could be a reduction in the capacitance of the bushing.
- A measurable parameter could be the leakage current – a variation indicating a change of capacitance.

The gathering of data and the analysis and subsequent diagnosis is not simple, even for relatively well known conditions. Returning to the medical analogy, symptoms of appendicitis are occasionally misdiagnosed with fatal results[5].

There is a diagnostic process which applies to humans, assets, and most other situations:

- Make measurements and observations (e.g. temperatures, pressures etc.)
- Perform analyses (consider levels, changes, trends, presence/absence)
- Identify symptoms.
- Perform a differential diagnosis. Where necessary gather more evidence.
- In parallel, consider contextual information (location, genetics, or design/manufacture details, known history, loading, number of operations etc.)
- Infer a root cause failure mode.

The process requires experience and expertise in order to be useful, with an understanding that the first inference may not be the actual answer required.

1.8.2 Understanding Variability

Measured values of data have inherent variability. Some reasons for variability are obvious. For example, a measurement of transformer power factor may be inaccurate: a SFRA result may be subject to series resistance offsetting the results, or a temperature gauge may be inherently imprecise or inaccurate. All measurements are subject to systematic or random errors, so even the most precise measurements made using the most accurate instruments contain some variability.

Even modern diagnostic tools can produce inaccurate results if they are not used properly or the necessary precautions are not taken. Measurements of the transformer bushings' power factors will

[5] From personal experience in a third world hospital; they did, however, observe exactly when the patient died.

potentially give very inaccurate results if the bushings are not properly cleaned to remove potential paths for leakage current. Those performing tests must understand the correct test procedure. For example, testers need to know the part of the asset that needs to be earthed/grounded, whether the test leads need to be shielded, how long the test voltage should be applied before a measurement is taken, and the process for taking an oil sample and storing the oil during transport to the lab.

In the past, many test tools were large, complex, and expensive. As a result, they were loaded onto test trucks which were operated by highly skilled testing technicians. Now very small and relatively inexpensive items of test equipment are available that can provide very accurate results. It is now possible for every technician to carry some of these test instruments in the back of their truck. However, unless the technician is properly trained and uses the test equipment frequently enough to develop the necessary skills and experience to properly use the instruments, the technician may be producing very accurate, but very incorrect data.

Highly accurate data is not always necessary to identify symptoms. For example, a doctor might simply place his hand on a patient's forehead to ascertain that the patient has a fever, as an exact measurement of temperature is not necessary in most cases; however, if the patient is being treated in a hospital, the patient's temperature might be taken every few hours. The doctor would be hoping to see the temperature trending back towards normal over time. In this case, the accuracy of the measurements is far more important.

Similarly, a hand placed on the radiators of a transformer might be adequate to detect a problem with the cooling system. If one radiator is cool whilst the others are too hot to touch, the radiator is probably not functioning correctly.

As highlighted by these examples, highly accurate measurements are not always required. It is, however, essential that the user of the measurements has some understanding of their accuracy and uses the data appropriately. Trending is a valuable tool when assessing the condition of assets, but useful trends cannot normally be established using estimates or highly inaccurate data.

Although inaccurate data can still be used for condition assessment (as long as the accuracy is understood), **incorrect data must not be used**. It is essential that those involved with both data collection and analysis look for obviously incorrect data. For example, if the oil temperature indicator of a heavily loaded transformer is reading less than the ambient temperature, there is almost certainly a problem with the temperature gauge. Similarly, if the winding hot spot of a transformer is cooler than the oil temperature, the data should be treated as suspect at best and wrong at worst.

There are many other examples of incorrect data. When the results from a dissolved gas analysis (DGA) show that all gasses are zero, the results are probably wrong. When the DGA results from an aged transformer suddenly show a drastic improvement, it is unlikely that the condition of the asset has changed; it is more likely that the oil has recently been changed or processed, oil samples have been mixed up, or the tests have been performed incorrectly.

Before analyzing any data, users should consider if the measurement is believable. To accomplish this, consider:

- The normal range for this type of measurement. If the results are all zero, or an order of magnitude higher than measurements taken from an asset in very poor condition, the data is probably wrong.
- Previous measurements, e.g., has the asset shown signs of deterioration since its last reading or has it miraculously improved?
- Other measurements and observations, e.g., if the asset has suffered a fault, and this is confirmed by other measurements, does this measurement also indicate that a fault has occurred?

It is possible that the data is correct and therefore is the symptom of an unusual failure mode, but users should do all they can to check the data before coming to this conclusion.

1.8.3 Timeliness of Data

Before analyzing any data, users need to understand when the data was recorded. Very accurate measurements that were taken five years ago will allow a good estimate of the condition that the asset was in at that moment in time, but not necessarily an accurate analysis of the current condition of the asset.

This information, however, may still be useful. In general, assets do not normally improve with age unless there has been some intervention, e.g., refurbishment, oil change, etc. Therefore, an assessment based on older data will probably result in an estimate of the best possible asset condition. Users must understand, however, that the asset may be in far worse condition than this assessment suggests.

1.8.4 Component Scores

Making one or more measurements allows the data to be analyzed and interpreted, the asset to be diagnosed, and a score for the measured component to be created.

Standards can be used to assist in the interpretation of the data. For example, IEEE C57.104 [16] can be used to interpret hydrogen levels in a transformer main tank. This is a simple process since the standard gives us condition codes 1 through 4, with 1 being normal or acceptable and higher scores indicating a worsening state.

However, there is no indication with those codes as to:

- What action, if any, should be taken, and
- How urgently the action needs to be carried out.

The limits that relate to each condition code are based on experience with a large number of transformers. The transformer that was measured may actually be designed, manufactured, or operated differently from most of the transformers from which the condition codes were derived. For example, a sample may be taken from a galvanized drain valve, in which case elevated levels of hydrogen may be present because of metal chemistry at the valve rather than because of a fault. Other variations based on specific designs are also noted within C57.104.

It is easy to misinterpret or misdiagnose the data and create the assumption that there is major fault or problem with the asset. Although this may cause panic for no good reason, a misinterpretation or misdiagnosis may also result in a lack of awareness of the true issue that is evolving. Problems could also be masked by this misinterpretation. For example, the elevated hydrogen levels resulting from the galvanized valve may be masking an evolving fault that is producing a lesser amount of hydrogen.

Competing interpretations for single measurement parameters, such as leakage reactance, or power factor or hydrogen levels, are possible. Competing interpretations of groups of parameters are also possible; different interpretations of DGA results would be a good example.

All of these issues, along with variability and timeliness of data as well as assessment of components, need to be understood and considered before attempting to relate the assessment of components to asset degradation.

1.8.5 Unavailable Data

It is common for some data or information that is needed for an assessment to be unavailable, particularly when large fleets are involved. This may be because a test was cancelled, or simply because data has been misplaced.

Users may still perform an asset assessment and determine a score by simply excluding from the assessment any failure modes that require unavailable data, then scaling the final score. Alternatively, users may estimate values for the unavailable data by using statistics or by using worst case or best case data to give a range on the final AHI.

Further discussion of these methods is beyond the scope of this document; however, it is important to note that where many assets have unavailable data, the accuracy of any resultant health index will be affected, further straining any relationship that exists between the health index and PoF.

1.8.6 Degradation

Blake [17] notes the relationships between degradation[6] and PoF; however, the measurement or measurements first need to be related to degradation. Note that in this report, deterioration is defined as a reduction in the physical value of some parameter, such as capacitance, while degradation is the loss of asset ability to perform its function. For example, what is deteriorating when ethylene levels rise in a transformer, or when the PD levels rise in a CT? At what point will degradation in performance be noted?

Often, in physical deterioration analyses, there is a well-established link between measurements and actual failures. For example, pipes, bearings, and steel frame members have been strained, stressed, or bent until they have failed whilst parameters have been measured. This allows correlations between the measured parameters and failures. Similar work has not been done to most large substation assets.

Blake also notes that models of degradation rely on many factors, including an understanding of location, load, manufacturer, fault history, inspections, and maintenance [17]. Assessing the degradation of a power system asset is complex and the knowledge of experts with appropriate expertise and judgment must be considered. Blake notes the different types of experts who are

[6] Although Blake has used the term "degradation," the authors believe that "deterioration" would be a more correct term.

involved in assessing assets, all of whom may bring an element of imprecision to the analysis of degradation/deterioration:

- Theorists: give options to select a model
- Consultants: actually select a model
- Asset engineers: determine parameters
- Maintenance engineers: assess asset condition
- Managers: decide how to use the results (interpretation)

The level of degradation that leads to failure depends on the various operational issues identified by Blake, and the physical details of the individual failure mode. A 0.50% power factor may be a strong indication of incipient failure in one transformer but not in another.

Time-based degradation in performance will affect different designs or manufacturer models in different ways. The development of computer-aided design to reduce margins in insulation, and the subsequent cost cutting exercises, have led to transformers which may be much younger but which are much more susceptible to the effects of heat and operational faults/switching. Transformers may be considered unique, hand-made chemical baths, and older transformers were often built bigger and more robustly than newer ones.

1.8.7 Working Forward

To work forward there needs to be definitive data which links measured parameters to the absolute level of deterioration in some operationally significant parameter, which is linked to an absolute likelihood of failure based on multiple and repeated tests to give statistical significance to the data.

Some work has been done in this area, but not specifically for power system assets and the operational context applicable to these assets. Paint thickness and performance on towers may be the most direct example of a possibly successful application of a parameter-degradation curve to power system assets; however, the link to subsequent failure, though monotonically related, is not linear.

It is possible to work forward to an asset health index by grouping (not necessarily weighting) parameter scores to yield a component score, and grouping component scores (not necessarily weighted) to yield an asset score; however, a meaning must be ascribed to the score. The meaning will depend entirely on how the score was derived. The problem we face is that the degree of uncertainty at each stage is compounded, such that the final health index is generally little more than a loose, non-monotonic, and non-linear ranking tool.

This will not help much if the number of assets to be ranked is small (such as at an individual generation station). The ranking may be little more than a Good / Bad classification to identify the poorer performers of the population.

As an analogy, consider the relationship between a car's tire pressure and the probability of the car failing. A lower than recommended pressure in a tire will be more likely to fail. Identifying a relationship is nonetheless difficult, and merging this relationship with all other parameters that could lead to a failure of the car, such as engine temperature, fluid levels, the pressure of the other tires, etc. results in a very complex problem. The process would be fraught with inaccuracy and imprecision at every stage. Furthermore, the relationship would be context specific. Failure is more likely to occur at high speed or during heavy braking, and less likely to occur when travelling slowly in heavy traffic. The relationship would also be dependent on the definition of failure, and would be different for each different model of car, and possibly even dependent on the day it came out of the factory.

To a degree, this type of approach has been attempted by the UK's regulator, OFGEM, in requiring the Distribution Network Operators (DNOs) to assess and rank their assets, both in terms of condition and risk. OFGEM has provided a framework to do this [18]. The approach has a lot of mathematical formulations and a number of assumptions and hypotheses. However, many of these assumptions and hypotheses in terms of the parameter values chosen and rates of failure expected. Regardless of this, the result is a consistent approach amongst the DNOs. It will take several years of feedback before the approach is completely verified or justified in practice. The approach recommended in the OFGEM document is discussed in Section 3.0.

1.8.8 Working Backward

Assuming the asset population under consideration is large enough that failure statistics are relatively constant over time, some progress can be made in terms of:

- Calculating the asset health index distribution over several years
- Identifying significant events which cause a change in asset health
- Identifying expected failure rates and the consequent numbers of failures expected
- Identifying candidates which are most likely to fail, based on *a priori* condition analysis and failure records
- Defining the various asset health index values to be associated with a certain probability or probability range
- Distributing assets to reflect expected performance
- Tracking prognostications to tune the system

The aim is to reduce population failures by identifying the units most likely to fail and improving the population performance as a result.

Published work, such as by CIGRE or by individual authors, provide examples of likely failure statistics for populations of different assets [3][4][5][7].

A review of Asset Health Indexing accuracy, based on forensic tear down of scrapped units in the UK, is discussed in Section 3.0.

1.8.9 Connecting Health Index to Probability of Failure

Calculating a PoF from an AHI based on pure analytics of measured parameters will be almost impossible, as the information base needed for individual parameters or for specific families to allow a physical model to be developed does not exist. Returning to the analogy of car tire pressure: when the pressure is 28 psi, what is the probability of failure of the car? What are the chances it will be able to drive another 1,000 kilometers?

Not all is lost, however. One approach to AHI is to define a set of categories or codes that cover a range of calculated AHIs. This approach is discussed in more detail in Sections 4.2 and 4.7. For example, set codes A through E as per Table 1-3 correspond to an AHI that ranges from 1 to 100; the categories do not have to be uniformly spaced or linearly related in time.

Table 1-3: Category AHI

Category	AHI Range	Description
A	0 - 30	Considered under normal operation
B	31- 50	Expect to replace within 15 years
C	51- 70	Expect to replace within 10 years
D	71 – 90	Expect to replace within 5 years
E	91 - 100	On replacement list for within 2 years

A notional PoF may be assigned in relation to the timescales associated with each category. For example, assume that a similar population of assets has a known failure rate of 0.5%. It could be assumed that a high failure rate would apply to the assets in category E, which are near the end of their life, and a much lower failure rate could apply to category A. The user could then estimate notional PoFs as follows in Table 1-4:

Table 1-4: Category AHI with First Estimate of Notional PoFs

Category	AHI Range	Description	First Estimate of Notional PoF
A	0 - 30	Considered under normal operation	0.05% PoF
B	31- 50	Expect to replace within 15 years	0.2% PoF
C	51- 70	Expect to replace within 10 years	0.4% PoF
D	71 – 90	Expect to replace within 5 years	0.8% PoF
E	91 - 100	On replacement list for within 2 years	1.5% PoF

Users need to multiply the number of assets in each category by the notional PoF to calculate the number of failures expected in that category, then add together the expected failures from all categories and divide by the number of assets to calculate the overall failure rate. If the overall failure rate does not equal 0.5% (the known failure rate for a similar population of assets), the notional failure rates need to be adjusted. This approach is discussed further in Section 5.0, where it is also noted that the base rate can be set to an unknown value, X, and multipliers can be used for each of the other categories, resulting in an approach to back-calculate X, the base rate.

Some users with access to failure and AHI data from large populations have used statistical information from their population to assign failure rates to AHI; however, even in a large population, the number of failures of assets with good AHIs would be too low to allow an accurate PoF to be calculated. Furthermore, many users do not allow assets assessed as being in very poor condition to operate until they fail—they try to avoid hard failures by admitting soft failures. As electric supply industry organizations tend to be risk averse, the result may be a higher proportion of soft failures than is valid. A forensic review is required to validate results, and such a review may indicate that some soft failures were in reasonable condition [7]. If units assessed as being in poor condition are removed from service and added to the failure statistics, the statistics become inaccurate unless the user conducts a forensic analysis of the asset to confirm that it was actually about to fail. Additionally, if assets assessed as being in poor condition are removed from service, but are not considered a failure, the statistics become distorted.

The expected failure rate for a population may be 0.5%. What is an acceptable failure probability for a unit on the system? Is it 1%, 2%, or maybe 5% in some cases? This depends on the individual organization and the consequence of failure. Deciding this failure probability is a risk decision, as Asset Managers may let some units fail while others are removed from service because they believe that they have an unacceptably high PoF.

1.8.10 Detailed Condition Assessment

If the AHI is based only on available data (i.e., it is not based on a full and detailed condition assessment), the AHI should only be used to identify assets that require further assessment. Before committing significant resources, time, and money to replace, repair, or refurbish an asset, a detailed assessment is used to justify the commitment.

Good condition assessment, however, requires appropriate measurements of appropriate conditions to determine the levels of degradation, deterioration, or incipient failure [5], [11].

Assessing against a population is relevant, but measurements must be used to determine condition instead of relying on statistical estimates for values that could be measured. Statistics apply to populations, usually suitably large populations. Statistics do not apply to individuals.

2.0 LITERATURE REVIEW

There is a lot of literature on asset health reviews and asset health indices, with many articles cross-referencing similar work. The drivers for the work are often regulatory or business and the approach taken is often to use health and criticality indexing to produce a risk ranking. The use of common Markov models occasionally drives a theoretical analysis, but there is precious little on linking, in practical terms, asset health to PoF based on physical experiments and true degradation processes [19]. There is, however, a lot of judgment and expert analysis.

This reliance on experience and heuristics is not necessarily a bad place to start, but closing the loop to check the validity of assumptions and findings has only been done a few times.

2.1 Theory and Practice

Hjartarson [2], [20] has presented papers that propose a system of component scores and weighted aggregation. The problem with such systems is that they dilute the effect of any one component or parameter (see Section 3.0 in this report). A number of organizations have followed this approach and ended up with scores between 1 and 100. The advantage of the approach is that it is easily understood. The disadvantage is that the relationship between measurable parameters and failure modes has been lost. A higher score does not necessarily mean a higher likelihood of failure, and an accurate sense of urgency is lost.

The example shown in Table 2-1 illustrates the weakness of the aggregated weighted system approach. Three transformers are given a health index based on the linearly weighted sum of nine components. In this example, a high score represents a bad assessment of the component. Scores range from 9 to 45, and are normalized to 100%. All nine components have been weighted equally for simplicity.

Table 2-1: Aggregated Weighted Systems

Factor	Score Range	Trf 1	Trf 2	Trf 3
DGA Main Tank Score	1-5	2	1	1
Dielectric Score	1-5	1	1	1
Thermal Score	1-5	2	1	1
Mechanical Score	1-5	3	4	1
Oil Score	1-5	1	1	1
DGA LTC Tank Score	1-5	3	1	5
Operational Score	1-5	2	3	3
Design/manufacturer Score	1-5	1	4	1
Subject Matter Expert Score	1-5	3	1	2
Sum	**9-45**	**18**	**17**	**16**
Normalized Sum (%)	**100**	**40**	**37.8**	**35.6**

Transformer 3 has the lowest score, and it is assumed that it is the least likely to fail; however, Transformer 3 has a score of 5 for the DGA LTC Tank Score, indicating that LTC is in very poor condition. With this urgent score, this transformer is actually the most likely to fail. Similarly, Transformer 1 has the highest score, and it is assumed, based on this scoring system, that it is the most likely to fail. However, the highest component score for any component of Transformer 1 is 3, so no component of the transformer would be highly likely to fail. Note that some users would add weightings to some components. For example, some users may decide that the score for the main tank DGA was very important and add a weighting factor of 2 to this score. In this example, this weighting factor will result in a worse score for Transformer 1, the transformer that is least likely to fail.

A discussion of such aggregated weighted systems is available [1]. Translating aggregated weighted systems into a PoF is not simple or direct since higher (or lower) scores do not represent a higher PoF (see Appendix D).

In the example illustrated by Table 2-1, the individual transformer scores may be very similar, but the timescales for intervention are not. It is obvious that Transformer 3, with a component score of 5, needs urgent intervention; however, Transformer 1, with the highest component score of 3, will most likely continue to operate satisfactorily for some time without intervention. The system is not calibrated and thus identification of a PoF is almost impossible. The calibration of scoring systems is discussed by Hubbard and is vital to making sense of the data [21], [22].

If the AHI is intended only to rank the assets based on overall or average condition, and it is not intended to relate the AHI to PoF, then a weighted system *may* be a good way to start. This system is relatively easy to understand, even if the dilution effects of the aggregated weighting rob the system of meaning; however, the result will not be directly relatable to PoF. Units at the top of the list may not be the units that fail.

The weighted aggregation system can be useful when used with other systems that better highlight the required time scales for intervention. For example, a very simple system that ranks assets according to their worst-case component score could be used in conjunction with a weighted aggregation. All assets that had one or more components with a very bad score, and therefore needed urgent intervention, would be ranked the highest. If the weighted aggregation score was shown next to the worst-case component scores, an asset manager could form an initial opinion about the overall condition of the assets that needed the most urgent intervention.

Tsimberg [23] discusses the well-understood concept of degradation being inversely related to PoF and notes that both degradation and PoF are inversely related to effective life. The author also presents the concept of a degradation curve that can change with maintenance activities. For example, circuit breaker maintenance can effectively reset the condition clock for many of the breaker's components. An Asset Health Index built on a combination of degradation with operational and contextual parameters relates well to Markov Chain and State Models of deterioration and degradation.

Zhou [24] gives a good review of the state of the art with relation to State-based or Markov models of asset health and modeled deterioration. The key is the need to take into account the censoring[7] of data available, as many assets of interest are still in operation, allowing for the increased accuracy of

[7] *Censoring* of data relates to the survival of experimental subjects through the time period of interest. Looking at failure rates of assets, any that have survived through to the present are said to be right censored, as the minimum age at failure is the present age, but the actual age at failure is still unknown.

the model. Similar work has been done building on basic Markov models [25], [26]. Such work is interesting, but most models discussed need to be extended to allow for multiple state transitions or direct transition from any state to a failed state.

Coullon [27] requires that an AHI represents not PoF, but residual asset life. This could be a good approach if the asset life is known on the day the asset was put into service. There is no direct relationship to PoF, but if an ageing model is assumed, PoF will rise with AHI since the AHI is related to remaining life. The relationship between failure and age, however, has not been well demonstrated, as discussed in [4]. As is shown in Section 5.7.2 of this report, random replacement of assets can generally reduce the average population AHI, precluding the validity of the residual asset life approach.

Coullon's use of an AHI is independent of his use of an Asset Maintenance Index (AMI). Coullon's AMI has similarities to some maintenance systems that derive a critical number for maintenance from combinations of time and condition or operations. The separation of a health index from a maintenance index is interesting and reflects utilities, such as National Grid UK, who do not include maintainable items in their transformer health index [7].

Dorison [28] describes an AHI as a weighted sum of influence factors and acknowledges that it is difficult to derive a PoF from an AHI. He notes the inability to derive a PoF is not a problem if the aim is to provide investment justification to regulators. The ability of the regulators to understand the pros and cons of any one system are not discussed.

Blake [17] assumes fault rates increase with AHI (where higher AHI values are worse conditions) and assumes a flat line roll off in degradation that produces a relation between measurable parameters and an AHI. To a degree, this is assuming the answer, as detailed in a CIGRE review of actual failure statistics [4].

Blake discusses a random or *non-condition* failure rate and an exponential rise in condition-based failures. Based on an initial status H_0 plus the random offset:

$$H = H_0 . e^{B(t-t_0)}$$

Figure 2-1: Blake's Equation

By choosing values for B, a score can be set to rise by a predetermined amount over a set period. Blake then discusses ways to vary the parameter B to reflect operational loading and other asset-based parameters. Blake also notes that experts can be biased in their assessments, specifically when addressing possible variations in the base equation and the parameter B; however, including their opinion makes their buy-in substantially easier.

DiMatteo [29] notes the need for a convincing derivation of an AHI in a business context. The article is based on big data, which may be loosely appropriate for some assets. DiMatteo describes a 1-100 score, which appears to be based on a weighted algorithm that includes age, maintenance, and DGA.

Edirisingh [30] echoes Blake's suggestion in regards to using expert input and possible sources of bias. Edirisingh also notes the need for an AHI to provide value to the user throughout an asset life, not just in terms of the operational piece: this leads to a multi-purpose AHI, which may not then provide value in any of its roles.

In an article focusing on process rather than on AHI in particular, Albrice [31] discusses the distinction between physical deterioration and obsolescence. Although this is not directly relevant to an AHI discussion, Albrice does note the inappropriateness of age when addressing condition statistics.

Cantler [32] focuses on cables and indicates that in a traditional AHI used to manage risk, age and type of cable were originally considered key factors. Based on his data and application, these parameters were no longer used in the AHI, but were replaced with condition measurements and evidence. Cantler uses a weighted system, so the standard caveats about dilution of parameters and inability to connect AHI to PoF to develop a risk index should be considered in conjunction with his work.

A Transpower New Zealand Review [33] is an example of a wish list of activities that should be considered when developing an AHI. The document lacks detail and specifics. It is possible that they will use an aggregated weighted index as a starting point.

Shepherd [34] gives one of the few non-utility based AHIs, the California Bridge Health Index (CBHI). The CBHI is used for a risk index for decisions on maintaining and replacing highway bridges. The work described in this paper seems to be driven by the need to create clear reports for individual California districts to use for investment justification. The method of using component level analysis with symptoms matches well to the utility systems presented in this document.

2.2 Closing the Loop

In a paper reviewing forensic tear down of condition-assessed transformers, DPT UK found a good correlation with many individual transformers [7]. Of 14 transformers identified for replacement, the review was completely correct about seven of them and partly correct about another four. Assessing solid insulation ageing correctly proved difficult in all three cases where the Asset Health Review was not correct.

This indicates that in 50% of cases, the result of the detailed health assessment that was carried out before deciding to scrap the transformers was not completely correct; and in over 20% of cases, the detailed review was incorrect. If this result is typical, the assessment done for a fleet screening AHI will most likely be less accurate. Users should be aware of this before drawing any conclusions from their assessments, an AHI based on their assessments, or from a PoF derived from their AHI.

2.3 AHI Discussion: What are the Issues?

Asset Health Indexing is relatively limited in use to the power utility world. This is surprising, and it may be because not all available literature has been reviewed, but currently few documents have been found that describe the use of AHIs outside this industry. Based on the literature reviewed, the most common reason for Health Indexing is to help address the unique regulatory requirements imposed on the power utility market; however, other markets do have regulatory requirements, so this still does not fully explain the main use of AHIs by the electric supply industry.

The literature review shows that, when talking with regulators, AHIs need to be clean, consistent, and simple. The AHI provides the tool needed to conduct a justification review. The capability of regulators and their staff to dissect and analyze an AHI is generally not discussed.
AHI formulas generally tend to fall into three categories:

1. Add up a number of figures that relate to the condition, age, or design of the asset. The result is a number that represents an arbitrary range from zero to infinity. This may include a number of common test results, but is likely to include other information that does not relate to the condition of any specific component such as age, loading history, family history, etc.
2. A weighted average of test results that represents degradation or failure that is then normalized from 0-100.
3. A weighted average of test results that are grouped into components and are then rolled-up using a normalized scale; usually 0-100. This method is by far the most popular.

Normalizing the score is very popular as it addresses the overall business need associated with the AHIs to provide simple comparison scores for regulators.

Several articles discuss the need to appreciate the differences between degradation and incipient faults/failures, both of which are folded into an AHI. Degradation is the long-term reduction in an asset's ability to function and may have more than one system or component involved. Usually these items cannot be maintained, refurbished, or replaced easily or economically. End of life is usually determined by such degradation.

3.0 SOME ISSUES WITH CATEGORY WEIGHTING

This section reviews data that is based on measurement parameters categorized on a 1-5 scale. Initially, the systems described have uniform weighting, and an interpretation of the results is attempted. Changes are then made to the data and conclusions are drawn.

3.1 Initial Data – Categories

15 parameters, denoted by letters A through O, are measured and categories 1 through 5 are assigned.

The timescales are not uniformly distributed among categories, as per Table 3-1.

Table 3-1: Category Labels, Timescale, and PoF Estimate

Category	Timescale	PoF Estimate
5	1 month	12%
4	1 year	4%
3	2 years	2%
2	5 years	1.5%
1	15 years	1%

Note the PoF value in Table 3-1 is difficult to justify and represents an annual equivalent. This brings in the need to identify an acceptable probability of failure. If an organization is happy with the present rate of failure, then that is an acceptable value; however, the rate may also be able to be improved. Along with weighting system issues, it is also important to examine the implications for interpreting a particular AHI as a source of a PoF.

In this analysis, if data is missing it is set to zero and the parameter is not included in the analysis.

3.2 Generating an AHI – Without Weights

This analysis includes 15 parameters, each with a category between 1 and 5, with a zero implying missing data. Each of the following statistics could be considered a candidate AHI:

- A simple sum of the scores
- An average of the scores – whether zero or not
- An average of the non-zero scores
- A sum of scores as a % of maximum possible score for all parameters
- A sum of scores as a % of maximum ONLY for those non-zero parameters

As an example, consider the data in Table 3-2. 15 parameter values have been translated into categories as per the limits in Table 3-1. Initial data is in the initial category column; some data have then been made unavailable (parameters A and I are now set to zero).

Table 3-2: Initial Parameter Data and Effect of Missing Data

Parameter	*Initial Category*	*Subsequent Category*
A	1	0
B	1	1
C	2	2
D	1	1
E	3	3
F	1	1
G	1	1
H	1	1
I	1	0
J	1	1
K	3	3
L	1	1
M	2	2
N	1	1
O	2	2
Sum	22	20
Average/All	1.47	1.33
Average/<>0	1.47	1.54
%/Max	29.3%	26.7%
%/<>0	29.3%	30.8%

The statistics for each parameter set in Table 3-2 are interesting—do any of them demonstrate robustness if they are used as an AHI? In the face of missing data, the following can happen:

- The sum has gone down when data goes missing, but the health has not improved.
- The average of all parameters, including the zeros, reduces; the average category is 1.47 reducing to 1.33, but health has not improved.
- The average of parameters which are present (<>0) has gone up as the missing data points were of the minimum category, but health has not improved.
- The % of maximum score based on all parameters has fallen, reflecting how the average works.
- The % of maximum of non-zero parameters has risen, reflecting the average of non-zero data.

Are any statistics a good indicator of need for intervention? The maximum score from the initial data gave a category of 3, an urgency of 2 years, and an annualized PoF of 2%. With some data missing, that maximum category has not changed.

The average categories do not change much, as they are still within the 1-2 range. The percent maximum does not relate to a timescale or a PoF directly. To get a maximum percent, every parameter must be at a maximum value, but this is unlikely.

When setting up an AHI, a clear strategy is needed to deal with missing data. If the two sets in Table 3-2 were from sister units, which one would provide most cause for concern? Which would require further data analysis and/or testing? A good AHI can start with very limited data and allow for general ranking and assessment.

What happens when there is no missing data, but parameters vary? The discussion as regards what the statistics mean continues and a standard deviation will be added to indicate the spread in the data. The initial set of data and four subsequent sets of data where parameters have deteriorated are shown in Table 3-3.

Table 3-3: Initial and Subsequent Data with Deterioration

Parameter	*Initial Data*	*Subsequent Data 1*	*Subsequent Data 2*	*Subsequent Data 3*	*Subsequent Data 4*
A	1	1	1	1	2
B	1	1	1	1	2
C	2	2	2	2	2
D	1	1	1	1	1
E	3	3	3	3	3
F	1	5	5	5	5
G	1	1	5	5	5
H	1	1	1	3	3
I	1	1	1	3	3
J	1	1	1	1	1
K	3	3	3	3	4
L	1	1	1	1	1
M	2	2	2	2	3
N	1	1	1	1	1
O	2	2	2	2	2
Sum	22	26	30	34	38
Average/All	1.47	1.73	2.00	2.27	2.53
Std Dev	0.74	1.16	1.41	1.39	1.36
%/Max	29.3%	34.7%	40.0%	45.3%	50.7%

The progress of data in Table 3-3 is as follows:

- The initial data is as it was previously, with appropriate statistics.
- Data 1: parameter F goes from 1 to 5—best to worst—and subsequently stays there.
- Data 2: parameter G goes from 1 to 5 and subsequently stays there.
- Data 3: parameters H and I go from 1 to 3 and subsequently stay there.
- Data 4: parameters A, B, K and M each deteriorate by 1 category.

What effect does all of this have on the statistics?

- The sum deteriorates by 4 at each step.

- The average category deteriorates by about 0.267 at each step.
- The standard deviation shows a wider distribution, then narrower as the scores all get higher.
- The % score increases by 5.3% at each step.

Each increase of a category by 1 step for any parameter supplies an increase in the percent score of 1.33%.

In each of the cases above, the subsequent data saw a change of category from 1-5, or two at 1-3, or 4 increasing one category. The number of individual category steps is the same in each case. In addition, it is true that the sum, the average, and the percent max all reflect an increase.

What the raw data means:

Subsequent Data 1:	Parameter F has gone from a 15-year timescale to intervention needed in 1 month, and a PoF estimated at 1% has gone to 12%. This is an urgent situation and requires a measured response; the changes in the AHI statistics do not reflect the urgency of the situation.
Subsequent Data 2:	There are now two parameters at most urgent – 12% PoF estimated; if these were independent parameters, their combined likelihood of causing a failure would increase to ~23.5%. The statistics yielded by the scores do not reflect the actual urgency.
Subsequent Data 3:	Two parameters going from 1-3 increase the statistics as per a 1-5 change; however, the urgency does not change greatly overall for the actual asset. The urgency is driven by the highest score and may be modulated by other elevated scores. In this case, two parameters changing from 1% to 2%, in the absence of any other change, would almost double the PoF.
Subsequent Data 4:	4 individual step changes of category, and the statistics reflect the same variation as the other subsequent data sets; however, the elevation in urgency is much lower, as the contribution from the four parameters rises from just over 5% to almost 9%.

In summary, the statistics used do not help identify an actual PoF based on the entries in Table 3-1: and the urgency of the situation. The AHI rises, reflecting a deterioration, but the effect is diluted through averaging.

3.3 Generating an AHI - Working Back

If the AHI is based on the %maximum, and two assets have dissimilar scores, is the lower score less urgent? Not necessarily—see Table 3-4. This table shows data from an original asset plus data from two sister units.

Table 3-4: Base Case and Two Related AHI

Parameter	*Base Case*	*Case 2*	*Case 3*
A	1	1	1
B	1	1	1
C	2	1	2
D	1	1	1
E	3	1	3
F	1	4	1
G	1	1	1
H	1	1	1
I	1	1	1
J	1	1	1
K	3	2	3
L	1	1	1
M	2	1	3
N	1	1	1
O	1	2	1
Sum	21	20	22
Average/All	1.40	1.33	1.47
Std Dev	0.74	0.82	0.83
%/Max	28.0%	26.7%	29.3%

When comparing Case 2 in Table 3-4 with the Base Case, the sum, average category and %Max have all reduced, yet the data indicates that the transformer is in a far more urgent state. A Code 4 from the assignations used in Table 3-1 implies a 4% PoF in the next year. When reviewing Case 3, the AHI is marginally deteriorated, reflecting the single step change in one category.

The problem? It is unknown whether a change in a statistic is due to one small change, or several small changes, or some combination of large/small changes.

3.4 Adding Weights

We can add weights to the scoring system to favor particular parameters. Table 3-5 uses weights that sum to 100 that then modulate the final AHI that has been based on the %Weighted Max. The weighted average gives a similar, proportional score. Table 3-5 has the same raw data as used in Table 3-4.

Table 3-5: Weighted Parameters for AHI

Parameter	*Weight*	*Base Case*	*Case 2*	*Case 3*
A	5	1	1	1
B	5	1	1	1
C	5	2	1	2
D	5	1	1	1
E	5	3	1	3
F	35	1	4	1
G	5	1	1	1
H	5	1	1	5
I	4	1	1	1
J	5	1	1	1
K	5	3	2	3
L	5	1	1	1
M	5	2	1	2
N	4	1	1	1
O	2	1	2	1
Sum	100	21	20	25
%/Max		26.0%	42.4%	30.0%

The results in Table 3-5 show the deterioration in Case 2, as the parameter weightings are tuned to reflect where the variation is.

The problem is, as in Case 3, the use of weights may not reflect an urgent case if the parameter is not strongly weighted. Parameter H in Case 3 has moved from best to worst—it is the most urgent and has a high PoF. Regardless, the AHI has not risen greatly. How can a small change in a key parameter and a large change in a minor parameter be differentiated? Without the raw data, the answer remains uncertain.

Adding weights will help clarify those situations where what is expected to happen actually occurs, but this method is much less effective where less expected events occur.

3.5 Extracting Meaning from an AHI

Can a PoF be deduced from an AHI as demonstrated in Table 3-5?

If the weights are uniform, then any category change of one step will produce a net change in %Maximum score. If the PoFs assigned in Table 3-4 are uniform (say 1%, 2%...5%) then a net change in the score can be linked to a net change in PoF; the link is not purely arithmetical in that probabilities can't just be added (see APPENDIX E). Simple addition could be close to a correct answer, depending on the PoFs assigned.

The left side of Table 3-6 assumes 5 categories and assigns linear PoF from 1% through 5%; the right side has linear PoF at 10% through 50%. On the left side, the 15 parameters are each assigned an initial PoF based on the lowest category, 1: best condition: 1%; the data is then changed by amending parameter O in column 1 "Large Step" to be the worst: category 5 or 5% PoF. Parameter L through O are then changed in column 4 "Small Steps" to be 1 category worse than the initial data, category 2 or 2%.

Table 3-6: Combining Assigned PoF

Parameter	Initial	1 Large Step	4 Small Steps	Parameter	Initial	1 Large Step	4 Small Steps
A	1.0%	1.0%	1.0%	A	5.0%	5.0%	5.0%
B	1.0%	1.0%	1.0%	B	5.0%	5.0%	5.0%
C	1.0%	1.0%	1.0%	C	5.0%	5.0%	5.0%
D	1.0%	1.0%	1.0%	D	5.0%	5.0%	5.0%
E	1.0%	1.0%	1.0%	E	5.0%	5.0%	5.0%
F	1.0%	1.0%	1.0%	F	5.0%	5.0%	5.0%
G	1.0%	1.0%	1.0%	G	5.0%	5.0%	5.0%
H	1.0%	1.0%	1.0%	H	5.0%	5.0%	5.0%
I	1.0%	1.0%	1.0%	I	5.0%	5.0%	5.0%
J	1.0%	1.0%	1.0%	J	5.0%	5.0%	5.0%
K	1.0%	1.0%	1.0%	K	5.0%	5.0%	5.0%
L	1.0%	1.0%	2.0%	L	5.0%	5.0%	10.0%
M	1.0%	1.0%	2.0%	M	5.0%	5.0%	10.0%
N	1.0%	1.0%	2.0%	N	5.0%	5.0%	10.0%
O	1.0%	5.0%	2.0%	O	5.0%	25.0%	10.0%
Overall:	13.99%	17.47%	17.42%	Overall:	53.67%	63.42%	62.68%
	Delta	3.47%	3.42%		Delta	9.75%	9.01%

For the left side of Table 3-6, the overall PoF for the 15 parameters can be calculated. 1 large step has a slightly larger impact than 4 smaller steps. The change or delta in PoF is similar for the two Steps.

For the right side of Table 3-6, there is a different set of PoFs for the categories, starting at 5% for category 1 and ending with 25% for category 5. The same process as for the left hand side of the table is followed. The resulting overall PoFs are much higher and the delta more exaggerated.

The tables show that combining PoF is not simple and leads to uncertainty. The tables also show that combining 15 parameters, each indicating a PoF of 1%, yields an overall PoF that is not just the arithmetic sum, but is still quite high: ~14%.

Should a transformer with an estimated PoF of 14% be left in service for a length of time? If so, for how long?

On the right side of the table, there is a transformer with a parameter indicating a PoF of 25%. Should a transformer with a 25% PoF be left in service? How about 63%? That is a business decision, but the answer may well be no.

The analysis of AHI meaning has been based on uniform linear weighting. If the analysis begins with low PoFs and a linear distribution across all categories, say 1% through 5%, then the change in %Maximum AHI score can be related to a delta in PoF and the margins of error will be small; however, the timescales for action assigned must also be linear. If the PoF is doubled, it would seem reasonable to halve the time within which action is taken. This is not the case in Table 3-4. The timescales are not linear and the PoFs are neither linear nor linearly related to the timescale, but the timescales may be reasonable for the decisions being made and for the generation of an AHI.

What happens when weights are added to the analysis? The link between the change in %Max and the net number of category changes is broken, as demonstrated in Table 3-4. This results in added focus for the AHI, but there is no means to relate the change in AHI to a change in PoF.

3.6 Summary

An AHI based on uniformly linearly weighted categorizations, which have a uniform and linear set of PoFs, does allow a relationship between a change in the AHI and a change in PoF. The PoFs assigned to each category must be very clear. The PoF should be linked to a timescale for action/intervention, which relates to the PoF. Care must be taken when setting up such an approach, as the combination of individual PoFs from multiple parameters is not simple arithmetic. Generally, a uniform linearly weighted set of categorizations, based on a linear assignation of PoFs to the category, allow for a monotonic AHI where a higher AHI has a higher PoF.

An AHI based on non-uniform weighted categorizations loses the tenuous link between AHI and original PoFs for each category and thus for each parameter. The monotonicity is also lost since a higher (or worse) AHI does not necessarily mean a more urgent situation.

Weighted or unweighted approaches to AHIs should be designed with caution and a clear aim in mind. When set up correctly they can give an indication of average condition. They usually do not relate to the urgency of response required.

4.0 EXAMPLES

The examples given here are almost all in current use for asset ranking. The link to probability of failure is not usually well defined, and may be difficult to justify or verify.

4.1 Delphic Approach

The Delphic[8] approach to AHI is based on expert opinion and consensus, and is built around experience and available data [1], [12]. Each asset of interest is assigned to an AHI category as shown in Table 4-1. This simple model is based on 4 categories, with 1 being Good and 4 being Bad.

Table 4-1: Four Category Delphi AHI

Code	Description
1	Asset is expected to last for the foreseeable future, and at least 15 years
2	Asset is expected to last up to 15 years but may need to be replaced in 5-15 years
3	Asset is expected to last up to 5 years and may need to be replaced in 2-5 years
4	Asset is on active list for replacement within 2 years

The approach uses condition information to rank assets and help plan for future interventions in a simple but justifiable manner. There is a timescale associated with each code. It can be used to check that the distribution of assets into categories is reasonable given known population failure rates. If the expected annual failure rate is 0.5%, then in the two years associated with code 4, it is expected that about 1% of the population will fail. If the number were higher, more would be expected to fail in the two years than in the populations that gave the expected failure rate.

Given that the units in code 4 are in poorer condition, they are expected to have a higher failure rate, and thus the 1% limit in the example here should be an upper limit. The Delphic approach is loosely defined, so the distribution of assets should be used as a guide rather than a constraint.

The Delphic system has been used in the US and other countries to rapidly generate an AHI for investment purposes in a regulatory environment. There are no formulae for deriving the AHI from raw data since this is left to expert opinion. There are also no formulae to compute a PoF given the AHI, although there is some implicit indication based on the timescales for each code, and the approach could be extended relatively easily.

4.2 Semi-Automated Delphi System

In this approach, an AHI is developed automatically based on available condition data, and then assets are assigned to condition codes manually, such as those shown in Table 4-1. Component scores are used in the automated generation of the AHI, and when data is difficult to analyze automatically, an expert scale is used to develop the component scores [35].

The AHI is built on logarithmic scales so poor scores stand out, as discussed in Section 3.0. A partial indication of the process is given in Figure 4-1, where transformer identifiers have been removed.

[8] Name based on the Oracle at Delphi who was said to predict the future.

Ratio	Rated P	Manufact	Design	sign	Year	Overall Condition: Now	Overall Condition: Mitigated	Possible I	Core and Windings: Dielectric	Core and Windings: Thermal	Core and Windings: Mechanic	Oil: Ageing	Oil: Contamina	OLTC	Exterior
400/132 kV	240 MVA	AEI Wythen	A04a	32	1965	[illegible]	[illegible]	8	100	100	1	13	10	3	10
275/132 kV	120 MVA	EEC	E11b	32	1959	[illegible]	[illegible]	68	30	60	1	130	10	10	10
400/275 kV	1000 MVA	GEC	[illegible]	[illegible]	1994	[illegible]	[illegible]	35	30	60		36	100		[illegible]
275/132 kV	120 MVA	EEC	E11b	32	1959	[illegible]	[illegible]	11	30	100	1	20	10	10	3
275/66 kV	180 MVA	CP	D07	12	1965	[illegible]	[illegible]	26	60	60	1	70	10	[illegible]	
275/132 kV	120 MVA	MVE	[illegible]	[illegible]	1957	[illegible]	94	57	30	60	1	160	10	3	10
400/275 kV	750 MVA	HHE	[illegible]	[illegible]	1971	[illegible]	[illegible]	47	3	60		140			3
275/33 kV	100 MVA	PPT	[illegible]	[illegible]	1972	[illegible]	[illegible]	5	[illegible]	3	100	13		[illegible]	10
400/132 kV	240 MVA	CAP	C04	32	1968	[illegible]	85	54	10	60	1	140	30	[illegible]	
275/132 kV	240 MVA	HHE	H07a	12	1964	[illegible]	[illegible]	26	1	100	1	70	10	3	
400/132 kV	240 MVA	AEI Wythen	[illegible]	[illegible]	1967	[illegible]	[illegible]	26	10	60	1	70	10	3	
400/132 kV	220 MVA	PPT	[illegible]	[illegible]	1967	[illegible]	[illegible]	24	1	60	1	63	10	[illegible]	10
275/132 kV	240 MVA	HHE	H07a	12	1964	[illegible]	[illegible]	23	1	100		63	10	[illegible]	
275/132 kV	120 MVA	EEC	[illegible]	[illegible]	1955	[illegible]	[illegible]	24	10	60	1	70		[illegible]	10
275/132 kV	240 MVA	HHE	H07a	12	1966	[illegible]	[illegible]	23	1	100		63	10	[illegible]	
275/132 kV	120 MVA	EEC	E11b	32	1959	[illegible]	[illegible]	22	30	60	3	43	30	10	
275/132 kV	120 MVA	FER	[illegible]	120	1956	[illegible]	[illegible]	19	3	60	1	50	10	[illegible]	
275/132 kV	180 MVA	FUL	[illegible]	[illegible]	1962	[illegible]	99	23	[illegible]	60		63	10	[illegible]	
400/132 kV	240 MVA	AEI Wythen	[illegible]	[illegible]	1967	[illegible]	96	26	[illegible]	60		70	10	3	
275/66 kV	120 MVA	AEI Rugby	[illegible]	3	1960	[illegible]	[illegible]	16	100	3	1	40	10	[illegible]	

Figure 4-1: Log Scale Semi-automated Analysis

The component scores include entries for:

- Design and manufacturer, which are manually entered based on known failure rates
- Core/windings based on data for dielectric, thermal, and mechanical performance
- Oil quality results
- Etc.

Each component may have several sub-components and each of these sub-components is scored on a logarithmic scale. The result is a league table based on condition. The table includes two scores for overall condition. The Now score is based on an assessment of the current condition. Options for mitigating problems, based on suspected failure modes and causes, are considered. The Mitigated score is an assessment of the condition that the asset would be in if mitigations were carried out. This allows the calculation of a Possible Improvement score. This is useful as it indicates whether a particular unit can be addressed successfully or should be removed from the system.

The league table is used as a basis for reviewing individual units and assigning them to one of a small number of categories, as per Table 4-2, with 1 being Bad and 4 being Good.

Table 4-2: Category Delphi AHI

Code	Description
1	In a state requiring replacement, expect to fail within 5 years
2a	Faulty transformer, expected to deteriorate to health index 1 within 5 years
2b	Faulty transformer, expected to deteriorate to health index 1 within 5-10 years
2c	Faulty transformer, not expected to deteriorate further
3	Transformer with known design defect but no active fault
4	No known fault or design defect

The approach has a number of interesting features, including:

- The different code 2s indicate not just how the transformer has currently been assessed, but also indicates future assessments. This is somewhat similar to Markov Models discussed in Section 2.0, but the smooth traverse of different states on the way from Good to Bad may not occur smoothly.
- There are timescales associated with deterioration, based on component scores and suspected failure modes.

This method was applied to a population of about 1000 transmission transformers. The process was as follows:

- Transformers from the league table were manually assigned a Condition Code.
- Codes were reviewed.
- Code 1 transformers were replaced or planned for replacement.
- The feedback loop was closed by analysis of units identified and taken off the system.

A review of units removed from service in 2011 and 2012 compared the as-found condition with the predicted Condition Code:

> A total of thirty transformers were scrapped during 2011 and 2012. The actual health index was the same as expected for twenty cases, differed by one category in five cases, and different by more than one category in five cases. The actual health index was better than expected in seven cases and worse than expected in three cases. [7]

This system has some strong benefits, since a condition code is justified based on data and agreed method of analyses. There is an implicit probability of failure in the timescales involved. Feedback analysis has shown that there is justification in the approach, but some room for improvement.

4.3 Exelon

Exelon [36] describe a long-standing Asset Health indexing system for transformers, breakers, switches, lines, cables, and batteries that supports maintenance prioritization and asset replacement. Exelon is working on modernizing the system to allow data to be uploaded automatically.
The system provides a ranking of assets, but no good indication of PoF.

4.4 SP Energy Networks

SP Energy Networks have a multi-purpose AHI [37] for any asset based on:

- Design Standards
- Deterioration
- Operational Issues
- Vicinity and Location
- Fault Rate
- Critical Issues
- Maintenance Spares

The actual calculations indicate that risk, instead of asset condition or health, is the driver within an Asset Management framework; however, only those issues that relate to Asset Health could be related to PoF. This system does not allow for the calculation of a PoF, as there is no simple way to connect the two in a meaningful and consistent way.

4.5 Hydro One

The work at HydroOne [37] aims to provide universal understanding and consequently acceptance of the health analysis and health indexing within the organization, and provides a common ground for discussion and investment strategies [38]. The actual AHI is a weighted system on a 1-100 scale. Dynamic and up-to-date data is preferred, and those data points that are not likely to trigger an intervention or investment are not considered.

Assets are categorized based on risk. Appropriate strategies for intervention are then defined.

As with other weighted aggregated systems for AHI, it is almost impossible to relate this AHI to a PoF.

In a separate document [39], HydroOne reviews their distribution assessment. There is a focus on high-value assets, which are ranked in a manner synonymous with AHI. There is also differentiation between defect management and deterioration. The system follows the methodology of a standard weighted aggregated AHI. HydroOne notes a need for calibration across asset classes.

It is stated in the HydroOne review (and in other papers and reports [20], [38], [39]), that a health index should have the following properties:

- Indicative: must indicate the overall assets' health.

- Objective: the index must be verifiable to industry standards, observations, and PoF.
- Simple: should be easy to use.

The problem with a weighted system is that the relationship between AHI and PoF is ill defined. Furthermore, as highlighted by Table 2-1, a weighted aggregated AHI may actually be masking assets in poor condition and may not be a good indication of overall health.

4.6 XCEL Energy Transformer Condition Assessment

XCEL Energy began a comprehensive Transformer Condition Assessment program in the early 2000s [40] with the aim being to prevent in-service failure. Their base criterion was to begin with the obvious and identify those units that were more deserving of deeper analysis. Their method involved:

- Comparing families,
- Looking for multiple problem indicators, and
- Looking at ways to provide feedback and adjust/improve.

The process allowed everybody to provide input in all stages, from review of available test data to ranking. Data was interpreted according to standards, experience, and relevance. Where data was found to be questionable, it was discarded.

The result is a manual coding system based on 1 being Good to 5 being Bad, with the distribution of assets in the classes being checked for quantity. Although there is no formal relationship identified between coding and PoF, it is accepted that that poorer units need to be addressed first.

4.7 TATA Power System

Kini *et al* [41] describe an approach to developing AHI for transformers. Some details of this approach are reproduced in Appendix C. The approach has clear aims, which are to:

- Organize available results and data;
- Manage transformers by ranking them to justify replacement, maintenance, etc.;
- Monitor the population over time; and
- Generate a dashboard for management.

In developing the Transformer Health Index (THI), measured parameters were subject to analysis and interventions planned by:

- Standardizing frequency of testing and limits for test parameters;
- Considering operational and maintenance data and using it for background context;
- Recommending (standard) correction actions to be taken in case of parameter violation;
- Generating violation reports at a centralized location; and
- Suggesting analysis for risk mitigation.

The THI developed is based on a 1-100 system, similar to others in the industry. Once sub-component weights are identified, major component analysis takes place as per Appendix C.

Figure 4-2 is a flowchart for THI. It shows that the process starts by identifying the analysis of basic or Tier 1 factors. Manually assessed factors that are described as Tier 2 are used to adjust the final THI.

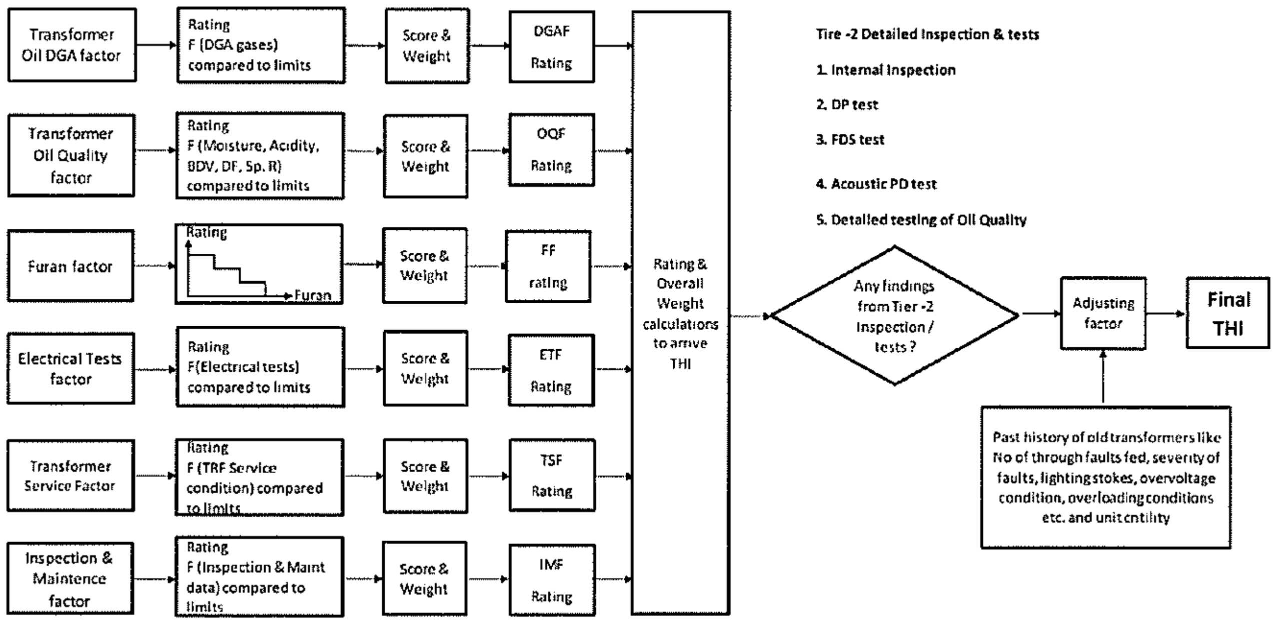


Figure 4-2: TATA Flowchart

Note that the Adjusting Factor includes expert assessment of operational data.

Health indices are given on a 1-100 scale, as shown in Figure 4-3. A qualitative assessment of the PoF is also provided with an estimate of the expected life.

Health Index	Condition	Probability of failure	Requirement	Expected life
86-100	Very Good	There is a low risk of failure and no additional maintenance or capital improvements are expected in the near term.	Normal Maintenance	More than 15 years
71-85	Acceptable	There is a low risk of failure and no additional maintenance or capital improvements are expected in the near term.	Normal Maintenance	More than 10 years
51-70		There is a medium risk of failure and detailed diagnostic testing / inspection will be required.	Increasing diagnostic testing	Upto 10 years
21-50	Poor	There is a high risk of failure and replacement or rewinding is required within the next five years to mitigate Imminent failure.	Plan replacement	Less than 3-5 years
0-20	Very Poor	There is a very high risk of failure and replacement or rewinding is required as soon as possible.	Immediately assess risk and replace	At End of Life (EOL)

Figure 4-3: TATA THI Evaluation

The process of developing a THI is formal, but there are areas where judgment and heuristics apply. The outcome is a weighted system focusing on known failure modes. The results have estimated timescales that provide a basis for a loose PoF calculation. The weighting approach means that the relation is not monotonic and thus difficult to calibrate and evaluate, and any sense of urgency is lost.

4.8 Austin Energy

At Doble's "Life of a Transformer" Seminar, 2015, Austin Energy [43] described a transformer Asset Health Index system, which they were using to support capital investment. The system has several levels:

- Initial Health: based on Manufacturer, Load factor, power factor, base DGA
- Dynamic Health: based on subsequent test data including further DGA and offline tests
- Current Health = Initial Health + Dynamic Health
- Criticality: based on system impact of failure
- Initial Risk = Current Health * Criticality
- Final Risk = (Initial Risk * Age Factor) + LTC Type Factor

The approach covers many aspects and uses weighted calculations and interpretations to derive the various scores. A lot of engineering judgment is used. The result is a league table, as per Figure 4-4.

NAME	Position	Equip. Class	Manufac-turer	Health	Criticality	Initial Risk	Final Risk	Ranking
Substation A	TR01	13KV	MFG A	141.7	1.2	170.04	210.04	1
Substation B	TR05	35KV	MFG B	71.7	1.9	136.23	203.04	2
Substation C	TR03	13KV	MFG C	131.7	1.2	158.04	193.84	3
Substation D	TR02	13KV	MFG A	76.7	2.1	161.07	181.07	4
Substation E	TR01	13KV	MFG B	56.7	1.35	76.55	164.21	5
Substation E	TR02	13KV	MFG D	36.7	1.55	56.89	142.58	6
Substation F	TR06	13KV	MFG A	59.7	1.8	107.46	138.21	7
Substation G	TR01	35KV	MFG B	56.7	1.3	73.71	137.4	8
Substation H	TR03	13KV	MFG B	71.7	1.55	111.14	136.7	9
Substation I	TR03	35KV	MFG E	41.7	1.2	50.04	130.04	10
Substation J	TR04	13KV	MFG D	83.3	1.55	129.12	129.12	11
Substation K	TR02	13KV	MFG D	54.7	1.35	73.85	128.62	12
Substation L	TR04	13KV	MFG E	41.7	1.55	64.64	127.87	13
Substation M	TR01	13KV	MFG F	46.7	2.05	95.74	125.31	14

Figure 4-4: Austin Energy Weighted Health and Risk Scoring

The Criticality factors seem to lie between about 1 and 2.1, and do not relate in any way to the health scores. The approach used at Austin Energy provides justification for an AHI and derives a value that is useful in determining which transformers are in poorest condition and which are at most risk. The system provides a basis for action and an opportunity to improve in the future.

The ability to determine a PoF is minimal, particularly since final scores are a function of criticality to the power system; but determining a PoF does not seem to be an aim for the system. The system, however, does work for its purpose, which is providing a ranking for investment.

4.9 US Utility

One US utility has shared some details of their AHI approach, using weightings to derive a normalized health score between 1 and 100 and adding a consequence score to allow for risk calculation; however, the system is set up such that if a bushing goes from the best possible to the worst possible score, indicating imminent failure, the AHI may not even cross the lowest level of concern.

To simplify matters, the AHI in this system has been renamed PoF; however, since this is a weighted system (which can mask imminent failures) and the final score includes a consequence score (which is not related to any failure modes), it is unlikely that this score can represent a PoF.

Units assessed as being in the high PoF category have not been represented in the actual failures that have occurred over recent years. Thus, the validity of the system is questionable.

4.10 Calpine

Calpine has applied a systematic analysis of their transformer fleet after two incidents in 2011 caused major interruptions. They have worked on a regular review with Doble Engineering [43].

Calpine uses a comprehensive review where available data is collated to produce an agreed scoring and AHI methodology, "setting standards and guidelines to continually evaluate the overall health of all oil filled transformers" [43].

The analysis is proprietary, but uses a weighted sum approach to identify long-term issues. Note that short term and operational response issues are dealt with separately during regular data reviews.

The result is a cost avoidance program that has demonstrated major savings. Costs avoided to date are in excess of $11M.

The aim of the analysis is to identify issues and plan appropriate intervention in order to reduce unexpected problems on their plant. At present, the aim is not necessarily to identify the remaining life of each fleet member, but that is a likely aim for the future.

The approach does not support the derivation of a PoF. The weighted scores are non-linear and provide only a general, non-monotonic relation to PoF. The system has, however, proven very useful.

4.11 UK Power Networks Report: Asset Stewardship Report, 2014

The UK Power Networks report is one of approximately 20 reports UK's Distributed Network Operators are required to produce for the UK Regulator [44]. This document is used to drive investment in the 132 kV network and to justify the replacement of existing infrastructure.

The 132 kV transformer age profile is shown in Figure 4-5.

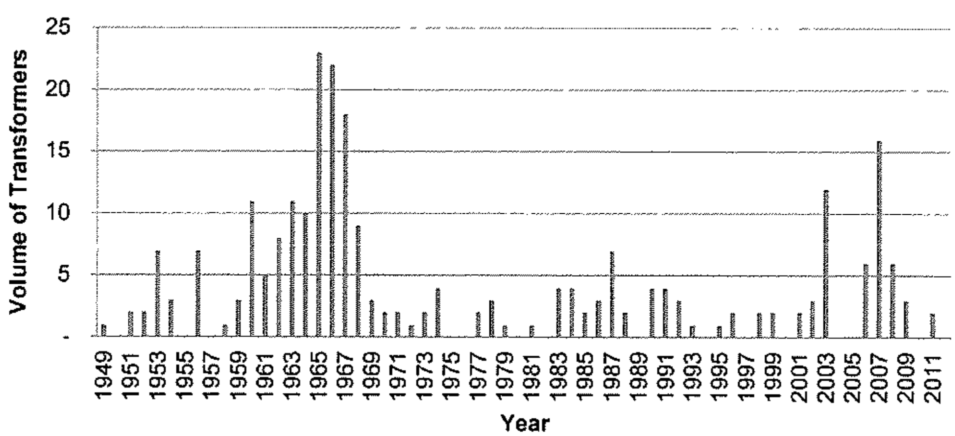


Figure 4-5: UKPN 132 kV Transformer Age Profile

The age profile shows a significant investment during the 1960s. The document notes that "although asset age is not a primary investment driver…it does have a cumulative effect on the serviceability of some of the assets" [44]. Reliability is linked to asset age only by cumulative operations.

Condition measurements are derived from inspection rounds, governed by the UK Power Networks (UKPN) document titled "The Substation Inspectors Handbook". The handbook details how to obtain data from handheld devices, maintenance results, and oil analysis [44].

The AHI is generated via a process of weights and analyses, including heuristics, as shown in Figure 4-6. UKPN notes that data validation and data completeness are issues.

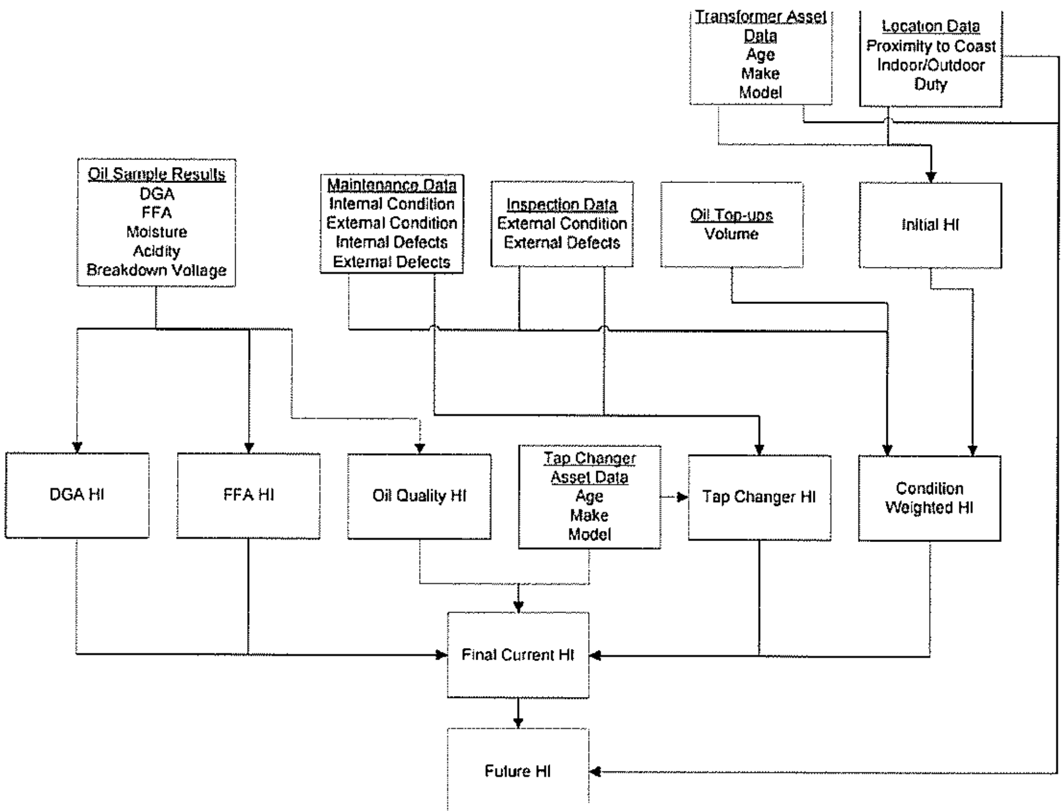


Figure 4-6: UKPN AHI Process Flowchart

The resulting AHI is a combination of several subsidiary AHIs and is not specifically linked to PoF; however, the AHI is justifiable in context and of value to the company.

4.12 Japan: CIGRE Poster

At CIGRE Paris, August 2016, Kobayashi et al. presented a poster on the latest work on AHI in Japan [45].

As investments in the face of deregulation rise to cope with increasing power demands, the authors note the challenges and benefits of an AHI as one of four key areas of focus. They also focus on:

- Consolidation of failure data,
- Predicted failure rates (with respect to economic life), and
- Substations as a complete system.

They developed an AHI for 66 kV transformers. It is a mix of condition and consequence, which, as a discussion with Mr. Kobayashi confirmed, makes it more suited to risk analysis. The potential problems are the intermingling of possible condition factors such as rising DGA or poor tap changer condition, with criticality factors such as the presence of PCBs in oil entailing a more substantial clean up. The system does not seem to be calibrated and has different scores possible for different contributing factors. *DGA* may contribute up to 10 points, while *winding configuration* may be 47.

The calibration of scales is not stated.

The system is proposed to allow for the ranking and identification of assets that need replacement. At present, the authors believe that there would be little chance of developing even a monotonic relationship between the AHI and PoF; however, further discussions with Mr. Kobayashi and his colleagues are planned to help develop a better understanding of the system.

4.13 ComEd Magazine Article

Graves, of Commonwealth Edison (ComEd), notes that their asset assessment initiative [46], beginning in 2007, weighted a number of components based on multiple data sources and resulted in a 1-100 score. The assessment was repeated quarterly to provide a snapshot of asset status. The present approach is to add in real time data, possibly including Phasor Management Unit data, to allow for more frequent updates and an analysis that is more comprehensive.

It seems that ComEd will continue with a weighted approach, and will use their AHI as an input to their risk analysis, which uses a traditional condition/consequence chart. There is no indication as to timescales for action, or for calibration of component scores and actions. The result is an inability to link AHI to PoF monotonically.

4.14 Kenya Power & Lighting Co. Transmission Power Lines

[47] shows the development of an AHI following a formulaic approach, featuring a weighted set of inputs and a 1-100 score. The authors assume a bathtub curve for asset ageing and failure rate. The key parameters are evaluated using expert analysis or reference values, and are weighted as per Table 4-3.

Table 4-3: KPL Transmission Line Evaluation

Condition	Grading	Weight	Max score
Corrosion on insulator hardware	1-5	6	30
Tensile test	1-5	3	15
Partial discharge test	1-5	4	20
Insulator data	1-5	6	30
Conductor general condition	1-5	2	10
Concrete hammer test	1-5	2	10
Contact Resistance	1-5	6	30
Infrared	1-5	6	30
Corona imaging	1-5	2	10
Acoustic test	1-5	2	10
Ultrasonic test	1-5	2	10
Tower general condition	1-5	4	20

PoF is discussed only qualitatively and the relation between the AHI and PoF is neither described nor identified.

4.15 KPL: Mombasa Network

In a 2012 paper, Bosire [48] describes KPL (Kenya Power & Lighting) work on transmission line AHIs with a focus on assets in one district. Four categories are defined, based on assessed imminence of failure and the previously-developed AHI:

- CR1 is a condition in which there is no detectable or measurable deterioration and no increased probability of failure.
- CR2 is where there is evidence of deterioration that is considered to be normal ageing and has no significant effect on the probability of failure.
- CR3 is a condition where there is significant deterioration that increases the probability of failure in the short to medium term.
- CR4 represents severe degradation and indicates an immediate, significant increase in the probability of failure.

The AHI is derived from a number of parameters, as shown in Figure 4-7.

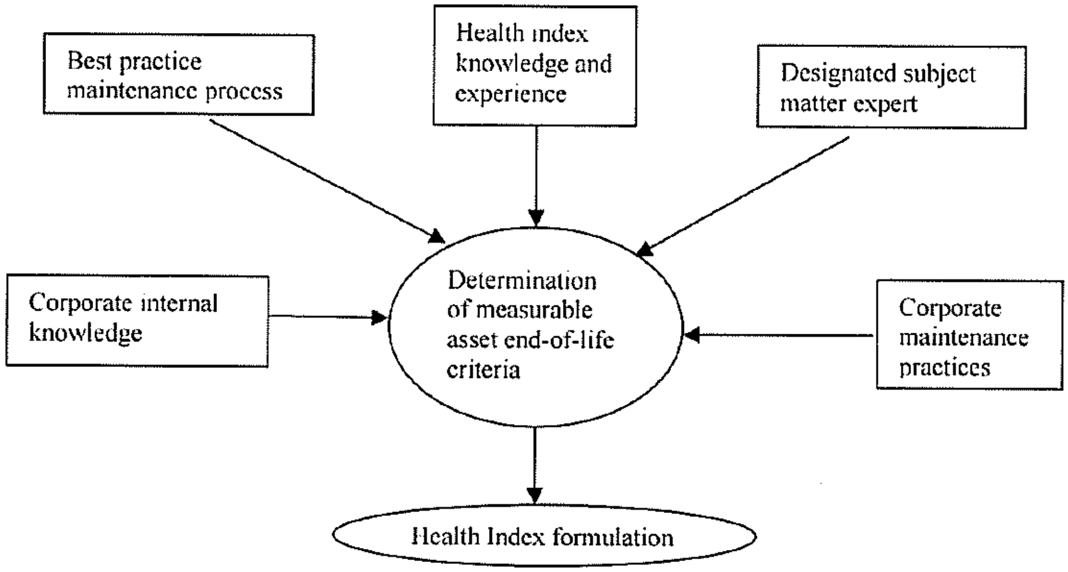


Figure 4-7: KPL AHI Derivation

The authors also look at the distribution of AHI scores and define further categories, as shown in Figure 4-8.

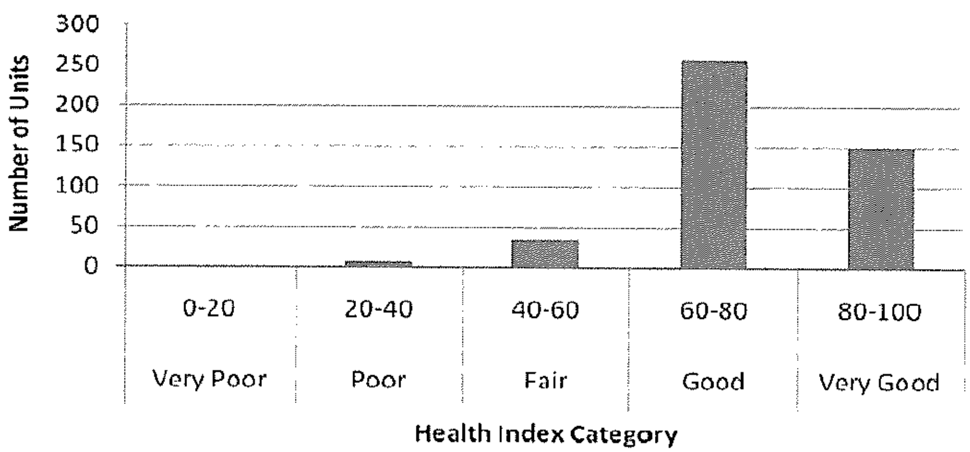


Figure 4-8: KPL AHI Distribution

A qualitative indication of PoF is also given, as per Figure 4-9.

Health Index	Ranked Condition	Probability of Failure (Pof)	Requirements
80-100	Very Good	Low	Normal maintenance
60-80	Good	Low but slightly increasing	Normal maintenance
40-60	Fair	Rapidly increasing but lower than pof at mean age	Increased diagnostic testing , possible remedial work
30-40	Poor	Higher than pof at mean age and increasing	Start plan to replace or considering risk and consequences of failure
0-20	Very poor	Very high and more than double the pof at mean age	Immediate assess risk, replace or rebuild based on assessment

Figure 4-9: KPL AHI and PoF

The relationship described is monotonic, indicating that the lower the AHI, the more likely a unit is to fail. The AHI is used to generate a PoF, and the PoF is used to drive action.

The problem is that weighted systems, as shown in Section 2.1, may not be monotonic, and some units with a higher AHI may actually have an increased likelihood of failure.

The work done here is heuristic, but does not give a true representation of PoF. In addition, the paper states that PoF will rise with age, which is contrary to what has been noted by many other investigations and analysis; however, this may be true depending on failure modes and the ability for condition to be improved via maintenance, e.g. painting.

4.16 UK Regulatory Approach for DNO Asset Indices

The UK Regulator for Gas and Electricity Markets (OFGEM) prescribes a system to calculate Asset Indices, including both health and probability of failure, for Distribution Network Operators (DNO) [18]. The document covers a range of asset types and has a range of tables and background data to enable translation of asset, location, and operation data into a health score, with subsequent modulation of the core based on observed and measured condition data.

A summary of the approach is given in Figure 4-10, where orange boxes indicate data input to the process by the user and blue boxes are values derived from the user data, look up tables, equations, and other parameter values prescribed by the system.

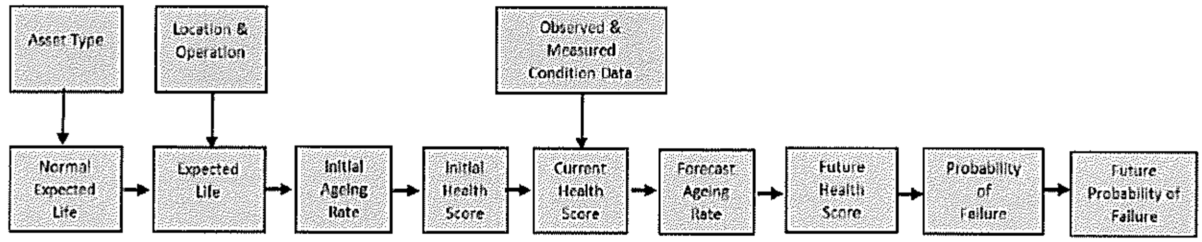


Figure 4-10: Overview of UK OFGEM DNO System

The method to derive the expected life of an asset is summarized in Figure 4-11. Orange boxes are used to indicate user input. Blue boxes indicate the output. Yellow boxes indicate how the look up tables are used with the input data and the given equations.

Given an asset's expected life, modulation by location and operation gives an actual expected life for a particular asset instance. The combination of Normal Expected Life and the subsequent modulation to give an Expected Life allows for the calculation of an asset-ageing rate.

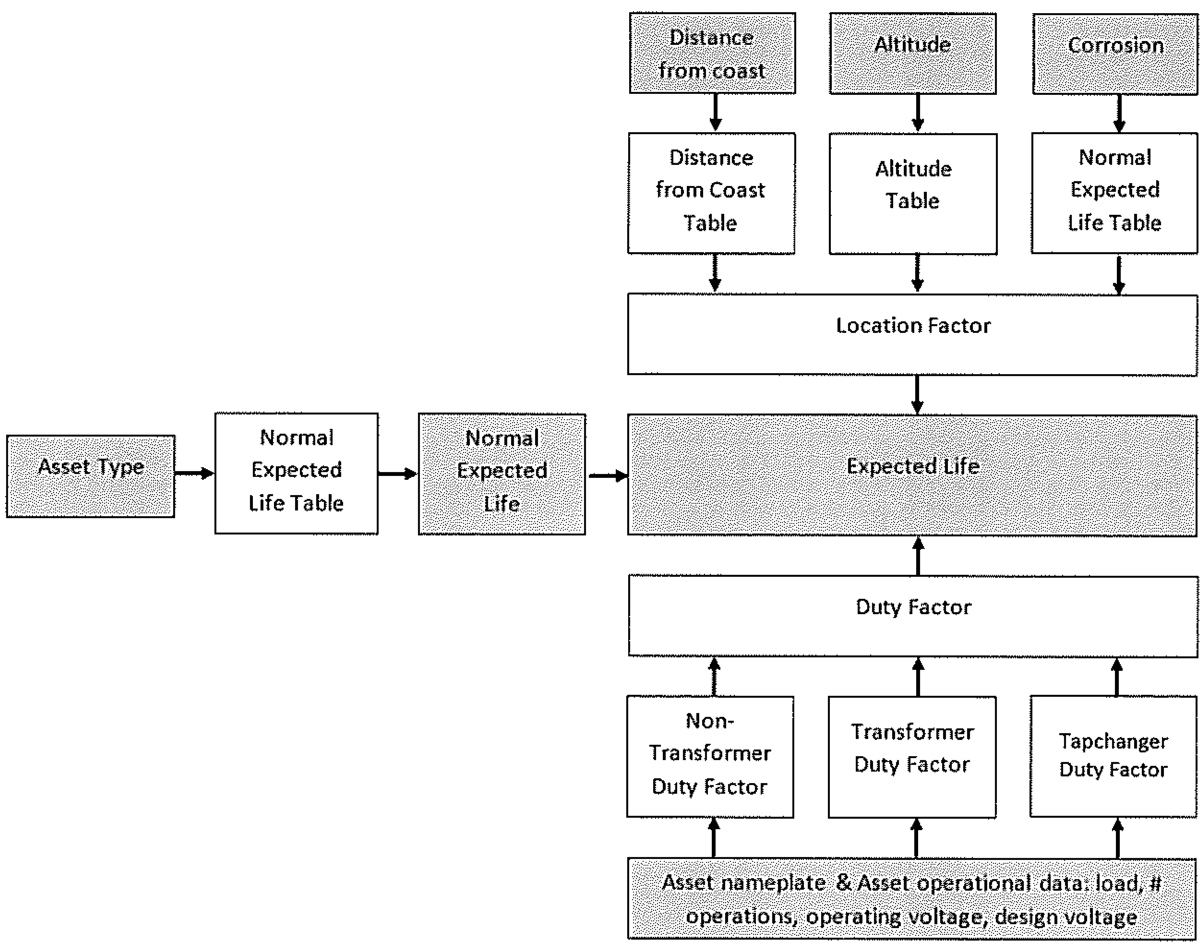


Figure 4-11: Derivation of Expected Life of an Asset

An initial health score is based on the asset age and its expected life. The score uses an exponential relationship between age and health and is capped at a score of 5.5. This value is then modulated based on observed and measured conditions and by asset-type-specific reliability modifiers. Future health scores are capped at 15. As noted in Section 5.0, age-based replacements can be self-justifying, even when replacing units at random.

Health scores are grouped into 5 Health Index Bands, as shown in Table 4-4. Average PoF values are used for all assets in each band, based on the numbers given in the OFGEM document.

Table 4-4: DNO Methodology HI Bands

Health Index Band	Health Index Banding Criteria	
	Lower Limit of Health Score	Upper Limit of Health Score
HI1	-	<4
HI2	≥ 4	<6
HI3	≥ 6	<7
HI4	≥ 7	<8
HI5	≥ 8	-

The derivation of PoF is shown in Figure 4-12.

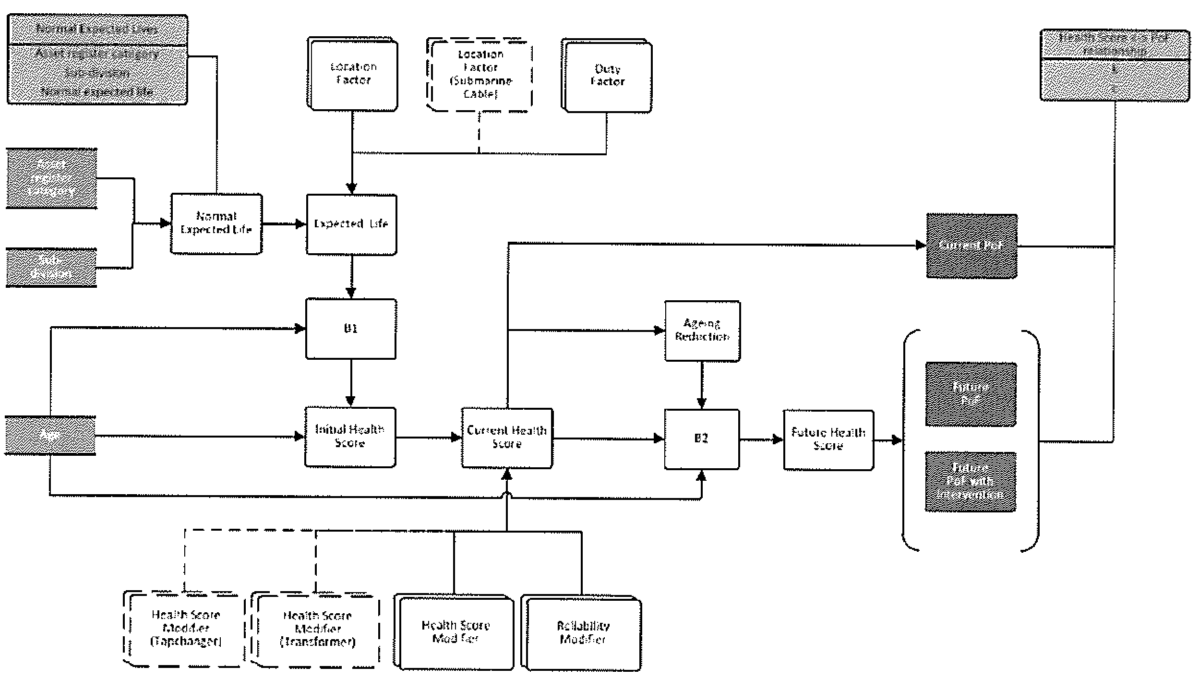


Figure 4-12: DNO Methodology Derivation of PoF

OFGEM's process is well documented and detailed. For example, the relation between Health (H) and PoF is given by the equation derived from the first three terms in a Taylor series expansion of an exponential (Figure 4-13). The function is assumed to describe increasing degradation, but not actually at an exponential rate.

$$PoF = K \times \left[1 + (C \times H) + \frac{(C \times H)^2}{2!} + \frac{(C \times H)^3}{3!}\right]$$

Figure 4-13: DNO Methodology Health and PoF Equation

K and C are constants to be determined. If H<4, then a value of H=4 is to be used. The K value scales the curve to desired limits. The C value is set so the probability of failure in the worst health state is 10 times that of a new or best health asset. It is assumed that there is no infant mortality and that failure rate only increases with age.

C and K values for different asset types are given. The expected asset life corresponds to H=5.5. The first significant signs of deterioration occur when H=5.5. The normal expected asset life is modified by location and duty factors, described in the document.

Initial and current ageing rates are calculated based on health score variation. Given that the health score is a function of age, it seems to be a circular calculation.

An ageing reduction factor generates a recalculated health score based on the present health score. This can be a way to reduce the offset and circularity introduced by the ageing rates and deal with inaccurate curve fitting.

The PoF of an asset is given as the maximum PoF of the sub-components. This is considered a reasonable approach, but one that does not properly take into account the laws of probability (see Appendices D and E).

Multiple measured condition inputs are used to derive a measured condition modifier. This is the point at which actual condition data is used. The Modifier acts as a multiplier on the age-based health score.

DGA is used to provide its own modifier, but it is based on 5 key gases: hydrogen, methane acetylene, ethylene, and ethane. Other gases are not recognized as indicating anything useful with respect to transformer condition. Scores are evaluated for each gas. The scores for hydrogen are given in Table 4-5.

Table 4-5: DNO Methodology for DGA Hydrogen Condition States

> Hydrogen (ppm)	<= Hydrogen (ppm)	Hydrogen Condition State
-0.01	20.00	0
20.00	40.00	2
40.00	100.00	4
100.00	200.00	10
200.00	10,000.00	16

The overall DGA modifier is based on the equation in Figure 4-14.

$$\text{DGA Score} = 50 \times \text{Hydrogen Score} + 30 \times \text{Methane Score} + 30 \times \text{Ethylene Score} + 30 \times \text{Ethane Score} + 120 \times \text{Acetylene Score}$$

Figure 4-14: DNO Methodology DGA Score

The percent change in DGA score is used to derive a DGA Test Factor from another calibration table. The approach does not use any standard interpretation schemes and does not identify failure modes.

Appendix D of the document discusses system failure modes, which are categorized into Incipient, Degraded, and Catastrophic. The document describes these failure modes in the following way:

> Likely Cost of Failure is the cost to return the asset to service (which may extend to full replacement of the asset). This is determined based on the three failure modes considered:
>
> - Incipient: The costs associated with addressing an Incipient Failure would not usually necessitate full asset replacement. Unless otherwise stated, a value equivalent to 10% of the Asset Replacement Costs has been adopted.
> - Degraded: The costs associated with addressing a Degraded Failure would not usually necessitate full asset replacement; however the works would normally be over and above those associated with addressing an Incipient Failure. Unless otherwise stated, a value equivalent to 25% of the Asset Replacement Costs has been adopted.
> - Catastrophic: A failure of this type would necessitate full asset replacement. Asset Replacement Costs have therefore been adopted, unless otherwise stated. [18]

These are really failure mode types rather than actual failure modes. Each asset type discussed in the document has components which may fail, which are grouped into the three types of failure modes. An HV transformer tap changer is classed as an incipient failure, while a bushing failure would be a degraded failure. This is interesting because actual failure modes are not discussed: not all tap changer failures are slow, and not all bushing failures are incipient. Either could be catastrophic in consequence. The approaches used in identifying failure modes from ISO would be relevant here (ISO 18095 [11]).

The relative proportions of failures for each type are used as a means to indicate the possible consequence of failure. 11kV and 20 KV transformers have a 60% probability of a failure being catastrophic, while for above 20 kV transformers the probability is set at 5%. Why the change? This is not discussed in the document. Failure modes do not seem to be defined, as would be required in a traditional Reliability Centered Maintenance (RCM) approach.

The DNO system may be seen as having a comprehensive set of inputs, analyses, calibrations and derivations of terms which include the words ***probability*** and ***failure***. This system will allow for a common methodology for asset assessment. In the opinion of the authors of this report, however, it will not actually calculate a monotonic AHI, and the results may be misleading as a consequence.

The resulting rankings are predominantly age based, the modifiers are possibly subjective, and the calibration tables are extensive, but there is little justification for any of the chosen values. Further, the results are not actually PoFs but modified AHIs which may not be monotonically related to the AHI. It appears that the relationship of Health (H) to PoF is tenuous and does not reflect standard failure rate models or experience. This is certainly the case for lower values of H.

The OFGEM DNO methodology is substantial: the document is 198 pages and covers a lot of different asset types. It does provide a way to view assets consistently across multiple organizations, which is somewhat beneficial in terms of making comparisons. The AHI derivations are very complex, are age skewed rather than condition based, and have intricate calibration processes. The lack of monotonicity means the final AHI loses direct meaning and any resulting actions are not calibrated by time. The complexity of the system may make the system difficult to implement effectively and the nature of the system may make it likely that users do not question the data, the system, or the results. This results in a "black box" approach that undermines the validity of the AHI process. An illusion of accuracy and precision may result, so use caution when applying these systems to ensure that interpretations are valid and meaningful.

4.17 Boston Water and Sewer Commission

Boston Water and Sewer Commission (BWSC) manages >800 miles of pipework, 600 miles of storm drains, and >80,000 structures [49]. An asset assessment was required to support a 25-year plan for system upgrades, capital improvements, and system maintenance. The work was based on a review of the previous 5 years' historical data, and an analysis to assess and rank assets.

The approach, called SCREAM (System Condition Risk Enhanced Assessment Model), requires all data to be coded according to predefined sets of possible conditions, with higher scores being worse. Instead of applying weights consistently across all components, the worst-case score for a component is taken as a base and other scores incrementally added. The codes are such that a score of 1 implies new condition while 100 implies failed condition. Grades are used to subdivide the 1-100 range: Grade 0 being Good and Grade 5 being Not Good.

Each component is given a defect score (DS) on a 1 – 100 base. For three components, A, B & C, the corresponding DS would be DS(A), DS(B), and DS(C). The overall score for the three components, where C has the highest component score, would be derived via:

Overall DS = DS(C) + ((100 - DS(C)) * (DS(A) *(Weight (A) + DS(B) * Weight (B))

Poor data led to misleading results when compared to newly generated inspection data, but these misleading results were used to drive further inspections. Figure 4-15 shows the results for pipeline grades for different installations.

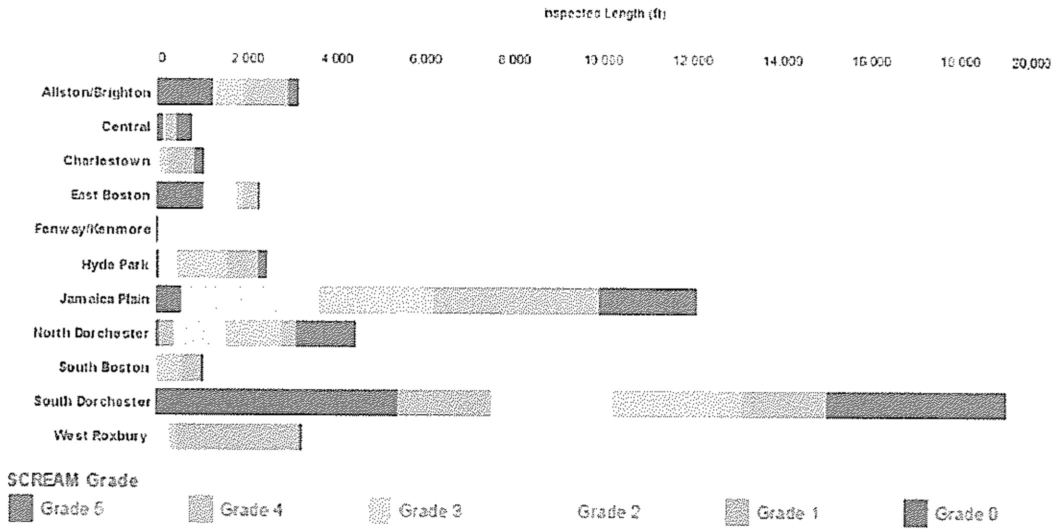


Figure 4-15: BWSC Pipe Length and Condition Evaluation

The BWSC approach allows for an auditable trail from raw data through to assessed condition and prioritised grading.

The scoring system retains an indication of the maximum score for a component. As with other averaging and weighted systems, using an inverse function to work back from the score to identify the maximum component is not possible.

The scores for a component relate to conditions that may not be calibrated in terms of time to failure. Thus, it may be impossible to relate the final grading score to an actual timescale; there is, however, some preservation of relative values, so higher scores are generally more urgent.

4.18 Small but Critical Stations

Staff from CERN made a presentation at "My Transfo 2016" conference in Turin, Italy which looked at prioritizing maintenance for their oil-filled transformers supplying the CERN particle accelerators [50]. They have a linear averaging system which looks for long-term health: it is based on six factors, each scored on a 1-3 basis, with an emphasis on ease of application by field technicians during inspections. The system also allows for assets to be allocated to a condition that requires urgent intervention should certain criteria be met, for example, in inspections or because of tests.

The system has several advantages in that it is easy to understand and apply; however, the system may provide misleading data in that an urgent score from an inspection that requires immediate intervention will be diluted by the averaging and not stand out in a final evaluation.

In discussion, CERN realize that their system requires further refinement as the competing needs for short-term intervention using maintenance and long term planning for replacement do not combine well in the linear averaging system.

4.19 Kinectrics IEEE DEIS Paper

In an IEEE publication [51], Kinectrics discusses their Health Index as a practical tool that combines the results of operating observations, field inspections, and site and laboratory testing to manage assets and prioritize investments in capital and maintenance plans. In an extensive paper, the authors describe how to categorize, weight, and collate data, including age.

The Kinectrics AHI is predicated on *remaining strength,* which is calculated from extensive weighted variables, and assumes that the remaining life will be related to overall strength but does not account for the dilution effect of individual failure modes through that weighting.

5.0 ASSET HEALTH INDEX AND PROBABILITY OF FAILURE

5.1 General Introduction

An AHI may be viewed as a means to capture disparate data in a simple form, whether to rank assets based on technical reasons or to justify the long-term investment plans of the financial group; it is unlikely that an AHI will contain surprises since it uses known algorithms to summarize well-understood data. Where there are surprises, they may relate to operational response being needed for sudden variation in a parameter rather than to the long term investment plan being readjusted by the millisecond as condition varies with load. Consequently, the AHI should be a concise summary of what is already known.

In generating an AHI, multiple data sources are included and multiple analyses and tools are used to end up with a single number. Therefore, much information is lost in the process[9]. This loss of data is well understood, and is acceptable if the result is useful. For example, producing an average and standard deviation from a set of numbers is a way to summarize the set, but does not have the same information content as the original set of numbers.

Before developing the AHI, it must first be determined what information should be gleaned from the one number. If that question is not clear, the value of the AHI will be limited. Encapsulation of knowledge into the one number may be precise, but if the answer is not related to the question, there may be an issue with any application of the results.

For long-term planning, a degree of imprecision is expected and is probably acceptable, but this degree should get smaller as the end date of the plan draws closer. For example, a plan for the replacement of transformers in 15 years may have a selection of 10 units. There may also be another 15 or 20 units in similar condition which are not expected to require replacement. As the replacement date gets closer, some units will have needed more attention than others, some will have already failed, and some that are not part of the replacement plan will have failed. Note that condition is only one reason for failure. External causes may also cause or contribute to failure (e.g., over-voltages, through faults, vandalism, or incorrect maintenance).

To generate a PoF from an AHI, the method of generating the AHI must first be examined. Is the function reversible? Are there timescales involved and inherent links from data to failure modes with known timescales of action? If not, when should the time element be added? Without the time element, a meaningful PoF cannot be developed.

This chapter will look at some of the uncertainty introduced in making measurements, performing analyses, collating data, and generating an AHI. Subsequent chapters will then look back at ways to generate a PoF from the AHI generated.

[9] Some AHIs may express the health of the assets in alternate forms, e.g., a colour or series of colours, a number and a colour, etc. They may contain more or less information than a single number, but some information is always lost.

5.2 Overview of AHI and PoF

An AHI takes a variety of data, often in continuous analog form, and generates a digital result. The conversion from analog to digital removes information; the recovering of that information is part of the process of generating a PoF.

Generally, we associate a higher PoF with a reduced residual life that requires more urgent remedial action[10].

If users want to translate an AHI into a PoF, the raw data and functions used to generate the AHI need to be known and understood in detail. In addition, the relationship between the input parameters is needed to see how they interact to produce an overall probability. The problem that many AHI systems face is that they can produce the same final AHI value in many different ways, so determining the urgency may be difficult. For example, an asset that is generally in very good condition but has a faulty component that urgently needs attention is given an average score in many systems. Another asset that is generally in a poorer condition but does not need urgent attention may be given a similar score.

If an AHI is ultimately intended to produce an estimate of the PoF, then the AHI must be related to expected failures and timescales. If qualitative values, which are not directly related to expected failures and timescales (e.g., low, medium, and high[11]), become enumerated as an AHI, then the resultant PoF is unlikely to be meaningful.

If a category or label of 1 through 5 is assigned for each input parameter or group of parameters, each category should have a well-defined meaning that is related to probability and timescale. For example, Category 5 may be defined as:

"Component in very poor condition. Repair or replacement required within 1 year."

When each of the assigned categories are weighted, collated, and used to generate a single number, any chance of a meaningful AHI may be lost if the functions are not well defined and related to probability and timescales. The sense of urgency may also be lost.

The effect of uncertainty, starting with individual parameters, should be considered, as the original measurement or data is not precise and covers a distribution. The uncertainty analysis can then be extended to the category assigned to the original parameter. Combining multiple categorized parameters with individual uncertainties may generate a lot of uncertainty in the final analysis, an uncertainty which is based on both the parametric uncertainty and the loss of information produced by the analysis itself (i.e., the functions used to combine data).

---

[10] Note that many assets require routine maintenance e.g. replacement of springs and contacts. Regardless of the assessed AHI or PoF, this routine maintenance "action" should be carried out as per the maintenance plan.

[11] If these terms are calibrated and related to expected failures and timescales, then a meaningful PoF may result.

Systems of weights can provide both simplification and a way to obscure the original data. The effect of weights is to dilute and average out significant changes[12].

A commonly used analogy for asset health and PoF is a car. What is the probability of failure based on engine temperature or tire pressure? Both parameters can be easily measured, categorized, and added together (possibly using weighting factors) to come up with a Car Health Index, but a meaningful PoF of the car is unlikely to be estimated from this index. An index that included additional parameters, considered the interaction of parameters where appropriate, then considered the probability of failure within a given timescale for each set of related parameters, may allow for a meaningful PoF to be estimated from the index.

5.3 Elements of an AHI to Be Considered

To derive a meaningful probability of failure from an AHI, the different elements that go into an AHI and subsequent PoF calculation, and the relationship between them, need to be understood and managed.

To derive a sensible result, parameters of interest need to be measured, variations recorded, failures observed, and their causes noted. This is a challenge because a statistically meaningful number of failures, which have been analyzed to accurately determine the interrelationship of the deteriorating parameters that resulted in the failure, are unlikely to be available.

Elements may be described as:

- **Raw data for a measurable parameter:** an indication of an acceptable range and the uncertainty in the measurement.
- **A view of the distribution of the parameter:** measured for different families or asset types. Unique, hand-made chemical baths are not commodities; they are not necessarily fungible, so anomalies or outliers in the raw data can be identified. When a parameter is measured, the uncertainty in the measurement at each level gives a distribution of likely values.
- **An indication of the relationship between the parameter and the probability of failure**: this is difficult. To do this properly, there needs to be a lot of transformer measurements and resulting failures so individual parameter effects can be quantified.
- **Using Anova table-style analysis reduces the number of measurements and indicates some parameter inter-relationships**, as with principal component analysis, but it still needs a lot of controlled experiments and measurements [52]. In addition, there is a need to look at historic values and try to deduce parameter-PoF relation from assumed logit or other function.
- **The AHI/PoF relation needs to be reasonably well defined in mathematical terms** to have value in collation to an asset PoF.
- **A categorization of data to group data into bands:** this is an Analog-to-Digital conversion, a broad quantization into discrete bands of a continuous quantity with analog measurement at near continuous values.

---

[12] Note that in some situations, the use of weighting factors might also highlight a significant change, if a high weighting factor is applied to a parameter that has changed.

- **Each category needs to be identified for bounds**: each category has an average PoF, a max and a min.
- **A translation of a parameter value into a category:** based on the uncertainty in the measurement, there is a percent probability that the reading is in category x, and a percent probability for each of the categories on either side.
- **Individual parameters yielding individual PoFs need to have a combination methodology** so that multiple parameter measurements can be combined to an overall PoF; Bayesian approaches and dependent/not-independent parameters.
- **Categorized parameters should be combined into an overall category** with a resultant PoF and degree of certainty.
- **Parameters may be grouped, categorized, and collated for a particular component.** Component scores are then collated for an overall score.

5.4 Parameter Measurement and Uncertainty

This section examines raw data and the means by which it can be related to true values: for example, a hydrogen level from an oil sample in ppm, or a temperature in degrees C. It also examines the uncertainty of this data and the impact the uncertainty has on its subsequent categorization, such as when the data is allocated to a category as a 3 when there are five possible categories.

5.4.1 Measurements and Distribution

When a value is measured, the measurement technique and measurement system provide both systematic and random errors. The result is only an estimate of the true value. Numerous measurements of the same value provide a distribution around the actual value. The distribution is a function of the accuracy of the measuring instruments and the repeatability of the measurements.

It is common for measurements to form a Normal or Gaussian distribution, symmetrical around a true value. An assumed distribution can be drawn based on the accuracy of the instrument and the confidence interval (C.I.) which indicates how likely it is for the true value to lie within the error range. The standard deviation is derived from the measurement value and error. C.I. is a measure of the spread of the distribution.

For example, in Figure 5-1, a measurement of a parameter is 25 units of measure, made with an accuracy or error of 10%. The accuracy of the measurement is usually supplied by the instrument manufacturer. Note that the vertical axis represents the probability that the corresponding value on the horizontal axis is the true or actual value.

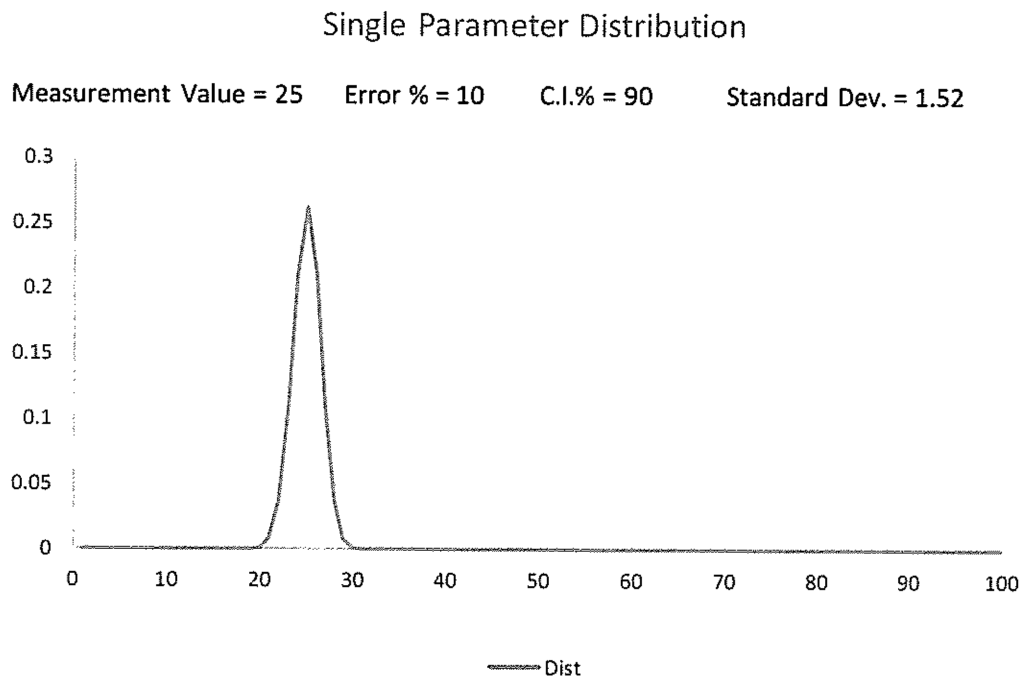


Figure 5-1: Parameter Measurement with Normal Distribution

For a Normal or Gaussian distribution, there are common values for the number of standard deviations within which the true measurement is likely to lie. For example, there is approximately a 95% probability that the true measurement is within 2 standard deviations and a 99.7% probability that it is within 3 standard deviations. The magnitude of the standard deviation depends on the magnitude of the measurement, the error, and the C.I.

The distribution of possible values measured for an individual measurement is different and independent from the distribution of the range of possible measurements over a larger population. For example, power factor measurements generally approximate an asymmetrical Poisson distribution, because the majority of the measurements from a population generally have a lower possible value, as shown in Figure 5-2.

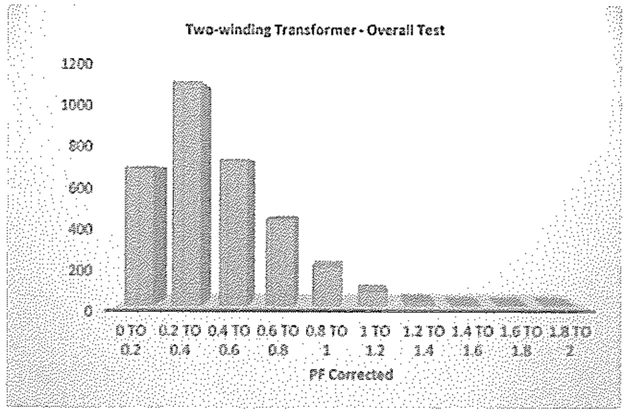


Figure 5-2: Distribution of a Population of Corrected Power Factor Results

The corrected power factor data for a large number of transformers, as shown in Figure 5-2, is predominantly below 1%. There is an extended tail to the higher values, which are less likely to appear in the population presented.

5.4.2 Probability of Failure

Although Figure 5-2 gives a distribution of results for a very useful condition assessment test, it does not indicate a probability of failure. Associating a higher power factor with a more deteriorated insulation capability, and thus an increased likelihood of failure, may be correct; however, this does not calculate or estimate the actual probability of failure. The proportion of the population which is above a certain value could be calculated, but that still does not yield a probability of failure. Further, the probability of a transformer failure would also be a function of other parameters and the interaction of some of these parameters.

To analyze probability of failure, a background in probability theory, Bayesian analysis, and dependent/independent variables, is required; see Appendix D.

The connection between measured parameter values and the probability of asset failure is difficult to make and is usually tenuous. For example, what is the relationship between tire pressure and the probability that a tire will fail? What, then, would be the probability that the car will fail, possibly catastrophically, requiring replacement? How does that probability of failure change if the tire pressure goes from 28 psi to 22 psi? Do these probabilities change as other parameters change (e.g., the speed at which the car is travelling)?

To have a meaningful function relating a measured parameter to a probability of failure, there needs to be data from multiple failures that would allow the relationship to be analyzed and a function derived with likely errors and uncertainties.

For example, with tire pressure, repeated experiments where pressures are recorded and then increased until failure is ultimately induced would need to be run. Other parameters such as ambient temperature, vibration level, etc. would need to be managed and controlled during these experiments. With a large enough sample of tires of the same design, a relationship between pressure and probability of failure for the tires tested could be identified; however, this is just one parameter that affects the tire, and the tire is only one of the many components that may result in the ultimate failure of the car. Furthermore, the analysis with one set of data is probably only valid for one design of tire. Other designs from the same manufacturer, or any designs from other manufacturers, would probably exhibit different failure characteristics.

For large assets, such as power transformers, the volume of data needed to relate measured parameters cleanly to a probability of failure for a particular design does not exist. It would be costly to induce failure in many transformers of manufacturer X and design Y while controlling all the other variables and changing only one measured parameter.

There is a possibility that using the Anova techniques developed by Fisher [52][13] can reduce the number of data points needed when compared to individual controlled analyses; however, failures covering a wide variety of parameters recorded across the range of independent variables would also need to be known and confirmed.

Given all of the problems described, what options are available to allow for a meaningful estimate of the PoF? One option is to estimate the curve that relates raw data to a probability of failure. For example, assuming a linear relationship between parameter values and probability of failure could result in a function such as is shown in Figure 5-3. The probability of failure has a maximum of 1.0, but it is likely to be very difficult to determine the value of the parameter at which failure is certain, i.e., a PoF of 1. The range of the parameter is scaled to show values between 0 and 10, and is plotted against a limited range of PoF values.

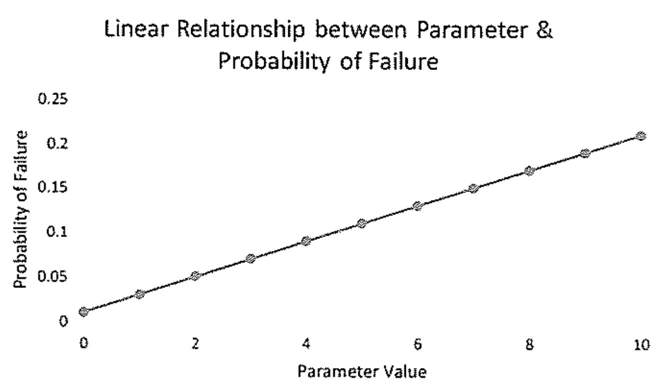


Figure 5-3: Linear Relationship for PoF

For some parameters the upper limit may be undefined, making a linear relationship unrealistic at high values. For example, a hydrogen value between 0 and 1000 may cover an appropriate range that can be assumed to correlate to a linear increase in PoF, but what happens if the level keeps rising? A logistic type function may reflect the relationship more appropriately, as shown in Figure 5-4.

[13] The reference does not specifically reference "Anova" tables or "Analysis of Variance" tables, but the details are clear.

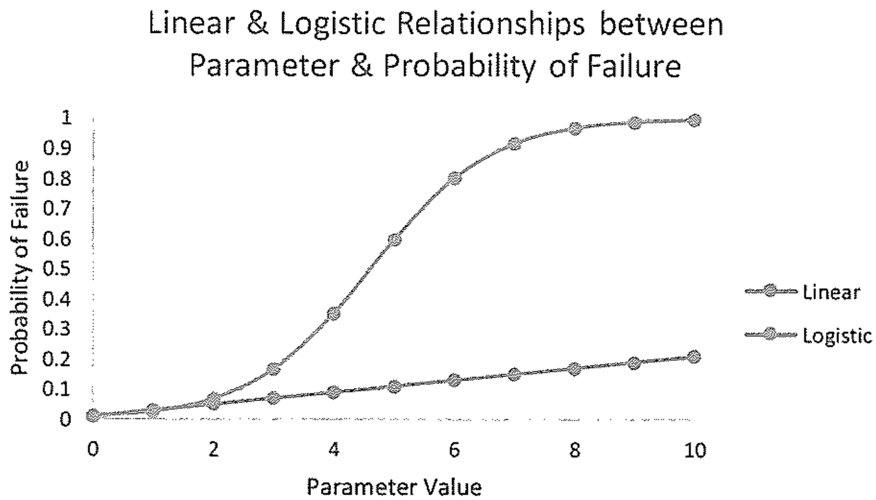


Figure 5-4: Linear & Logistic Relationships

The logistic relationship, as shown in Figure 5-4, is one of several such curves that can be tuned to reflect the limited measured data available. Bill Bartley, then of Hartford Steam Boiler Insurance, used such curves to model likely population failure rates in general in analyses of insurance statistics [58]. The upper limit of the logistic function value asymptotically approaches 1, reflecting the unlikelihood of a failure being certain. The lower, minimum probability value can be set to an appropriate level, reflecting random or externally caused failures. Both the point where the curve flexes and the rate of change can also be set.

The curves, shown in Figure 5-4, are examples that could be used, but they do not show the uncertainty or the error bounds. With the scarcity of actual parameter-failure data, such curves are usually ill-defined and usually shown without errors or uncertainty. This can be misleading as the errors are so large that three standard deviations could fill the chart.

5.4.3 CIGRE Brochure 296 DGA Data

The CIGRE Technical Brochure 296 [53] summarizes data from oil samples taken from the bottom of the main tank of a population of transformers. The samples were taken "shortly before or after" a fault, so they can be assumed to correspond to the fault. The timescale of "shortly" is not defined.

IEC 60599 is referenced throughout the document and is used as the basis for the analysis of the DGA data. Only three companies provided data for the exercise detailed in TB 296. Consequently, only a very small amount of data is used to develop charts that link gas concentration to probability of failure. Any company wishing to perform a similar analysis can utilize IEC 60599 (or equivalent IEEE C57.104) and use their own data. Results will depend on specific equipment and operating conditions.

TB 296 defines the probability of failure in service (PFS) as:

PFS = (# cases with high DGA before an event) / (total # of analyses at any level of DGA)

This definition results in an indication of probability based on a test, but not in an indication of the accuracy of the test, e.g., false positives, true negatives, etc. Consequently, this is not an ideal definition or method for calculating PoF. Please refer to Appendix D: "Some Aspects of Probability Theory."

For the three companies providing data, the Pre Failure Gas Concentrations (PFGC) are defined as the level of the top 1% of results. For acetylene, the three companies have PFGC between 300 and 600 ppm. These figures are high. Figure 5-5 is the chart linking actual ppm, the normalized percent, and the probability of failure (PFS).

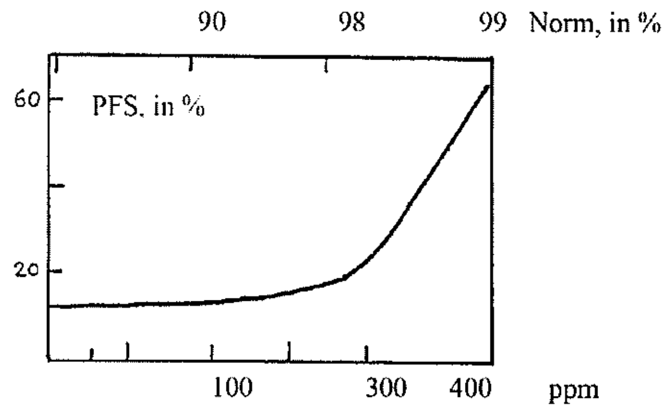


Figure 5-5: Probability of Failure and Acetylene Concentration (Copyright CIGRE 2006)

Figure 5-5 shows the sharp rise in PFS which might be expected from a logistic function. This is reasonable as an approach, but to define the PFS based on a small population needs an investigation of the raw data to determine the uncertainties. Usually a DGA's values are within 10-20% of the true value of dissolved gasses, so there may be significant variability when there is a distribution.

The chart is also extended to low concentrations. For a DGA sample with virtually zero acetylene, the PFS is around 15%. Is that a valid result? Are there about one in six units with low acetylene failing each year? Is a year the appropriate timescale? (The timescale is not defined.) The probability of transformers failing with a low ppm level of acetylene is very difficult to deduce, since most records of failures relate to high levels of acetylene.

The TB 296 analysis would also benefit from a Bayesian inference, where probabilities are updated as more evidence of information becomes available. Furthermore, the following should be considered when reviewing the information in TB 296:

- That the defined value for PFS is based on values before a failure; the number of units that had similar gas values but which did not fail should also be known.
- The prevalence of failure in the population.
- That the present system assumes a uniform distribution of oil samples being taken across the population. More samples taken could drastically lower the PFS, but the provided data does not confirm this.
- All of the conditional probabilities would need to be known or estimated before the conditional probability of failure given a certain dissolved gas level could be calculated.

– How big are the data sets? Were any data sets left out?

The CIGRE TB 296 is interesting and may provide an indication of a limit above which a failure is more likely than not. The actual ppm levels are high and are based on an indeterminate amount of data. The resulting chart has a relationship, but this cannot be verified because the raw data is not available. The implication that very low concentrations of acetylene could have a PFS of 15% seems excessive.

The aim of this discussion is not to criticize the efforts of the CIGRE experts who contributed to this brochure over 10 years ago, but to show how difficult it is to relate parameter values to an actual probability of failure. The CIGRE work is a good base for discussion, but shows that the need to identify appropriate parameters for the individual population and control parameters of manufacturer/design and operating conditions is not a significant consideration. This is highlighted as the three companies that provided data have significant variation between them in terms of the ppm levels of note.

5.4.4 What Are Acceptable Probabilities of Failure?

What is an acceptable probability of failure? How accurate must this value be?

If present failure rates around the world for an asset vary between 0.2% and 2% for different organizations and asset types, a realistic rate would fall between those two values, but an acceptable level also depends on the consequences of failure and the business context.

The consequence of failure is a function of the size and type of the load and a function of the redundancy of the system at the point of supply. Safety, environmental, economic, and other issues also need to be considered.

5.5 Parameter Categorization—Including Uncertainty

One approach to relate parameter values to PoF is to categorize the parameter into bands, with each clearly defined band being numbered/labeled/named. It can generally be assumed that assets with parameters in the bands that represent a worse condition, are more likely to fail. Categorization is the compartmentalization of parameter ranges and associating the compartments or categories with deteriorating condition.

Using a measurement of dissolved gas levels, 4 distinct categories based on IEEE C57.104 condition codes could be defined (as shown in Table 5-1).

Table 5-1: IEEE C57.104 DGA Condition Codes

Status	Dissolved key gas concentration limits [μL/L (ppm)[a]]							
	Hydrogen (H_2)	Methane (CH_4)	Acetylene (C_2H_2)	Ethylene (C_2H_4)	Ethane (C_2H_6)	Carbon monoxide (CO)	Carbon dioxide (CO_2)	TDCG[b]
Condition 1	100	120	1	50	65	350	2 500	720
Condition 2	101–700	121–400	2–9	51–100	66–100	351–570	2 500–4 000	721–1920
Condition 3	701–1800	401–1000	10–35	101–200	101–150	571–1400	4 001–10 000	1921–4630
Condition 4	>1800	>1000	>35	>200	>150	>1400	>10 000	>4630

Note that the codes in Table 5-1 do not indicate an action, a timescale for action, or a probability of failure. The codes just relate to higher DGA values and the assumption that higher DGA values relate to a more deteriorated condition. An assumption of relative probability (i.e., that higher codes have a higher probability of failure) could be made; however, it is unknown how much higher the probability is for each condition code.

It is not necessarily straightforward to decide how many categories to create and the boundaries for each, but this is still an integral step. It is also important to ensure that there is consistency between categories: a particular category for a hydrogen reading should have a similar implication for the action required and the timescale as the same category for carbon monoxide. Each category label should have a consistent timescale for action to indicate the same urgency of the results.

A set of category bounds can be overlaid on Figure 5-1. In this case, 5 categories have been chosen which are uniformly distributed between parameter values of 0-100. The data table for categorization is shown in Table 5-2.

Table 5-2: Initial Category Bands for Data in Figure 5-1

Category	Lower Limit	Upper Limit
1	0	20
2	20	40
3	40	60
4	60	80
5	80	100

There is no requirement to make the bounds uniform or to limit the number of categories to 5.

Figure 5-6 shows the categories overlaid onto Figure 5-1. The original measurement has an uncertainty that places the reading in the second category 99.95% of the time. It would be in a lower category just 0.05% of the time.

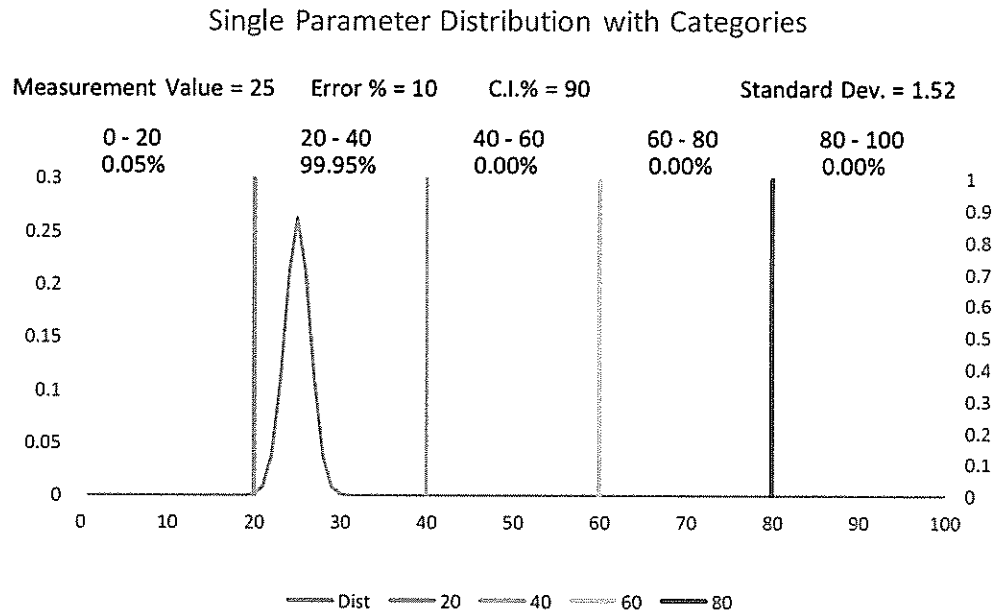


Figure 5-6: Five Categories Overlaid on Measured Parameter from Figure 5-1

If the reading is now changed to 41, and it is assumed that the percent error and the C.I. of 90% does not change, the result is now a standard deviation of almost 2.5 and the likely spread of readings is higher, as shown in Figure 5-7.

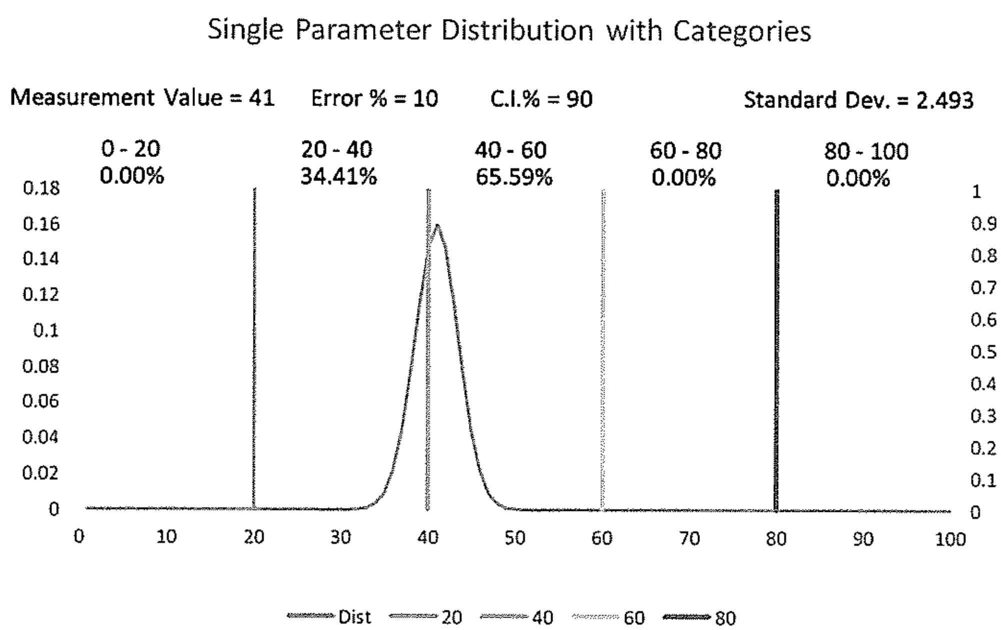


Figure 5-7: Five Categories Overlaid on New Value of Measured Parameter

The fact that there is a measurement close to the category boundary at 40 means that a significant proportion of likely true readings fall in the category below where the measured value lies: 34.41% or just over a third.

The less accurate the measurement or the less precise the C.I., the wider the spread of possible true measurements. This increases the chance of the consequent assignment of the score to the incorrect category.

One of the consequences of categorization is that the original raw data may move out of sight. In the charts shown here, a reading of 25 is very likely a category 2 (almost 100% certainty), and a reading of 41 is likely to be category 3 (66% certainty) but may also be a category 2 (34% certainty). If AHI calculations only use the final category, the raw data (the function used to categorize and the uncertainty resultant in that categorization) moves out of sight. Subsequent calculations have an unknown degree of uncertainty.

Any labels can be chosen for the defined categories. Numeric labels make calculation of a simple asset health score easy. As discussed earlier, simple numeric scores can easily hide problems that need to be addressed urgently. Simple numeric scores can also be used with weighted calculations, but weighted calculations can further obfuscate the raw data and its meaning. Non-linear or log scales can be used to assign numeric labels. The advantage of a log scale is that it makes more urgent data stand out in a way which linear data does not. Labels can also be formed from letters, which can be easily ordered, but prevent the simple sum of the labels assigned to the parameters being considered for the asset.

5.6 Multiple Parameter Combination

This section examines the means to collating categories from more than one parameter into a final score. Some of the techniques here were discussed in Section 2.0. Without a clear route to the raw data, there is little chance of producing a meaningful PoF from an AHI, but it may still be possible to generate a relative value. It may, however, be impossible to generate a relative value from many weighted systems.

5.6.1 Dealing with Uncertainty in Multiple Parameter Categorization

The data charts and categorization in Section 5.5 were for a simple case of a well-bound parameter, ranging from 0-100, with evenly distributed categories. If there were two parameters, the resulting analyses would be similar, but the collation of resulting categories would be more complex.

Table 5-3 features two parameters: Hydrogen in ppm and Temperature in degrees C (degC). Both measurements are given as 95.00, which may be inappropriate precision.

Table 5-3: Two Parameter Reading and Statistics

	Hydrogen	Temperature
Measured Value	95.00	95.00
Error +/- %	10%	5%
Lower Limit (-%)	85.50	90.25
Upper Limit (+%)	104.50	99.75
C.I.	95%	95%
# of Std dev (from C.I.)	1.96	1.96
Std dev (ppm or degC)	4.85	2.42

In Table 5-3, the error in the measurement is given as a percent, allowing a lower and an upper limit for error calculations. The Confidence Interval (C.I.) is given as a percent, allowing the number of standard deviations covered by the error to be calculated and the value of the standard deviation to be derived for each parameter.

Using the same approach as for the single basic parameter, boundaries for 5 categories can be assigned, as shown in Table 5-4.

Table 5-4: Limits for Two Parameters—Hydrogen and Temperature

Category Range		Parameter	Hydrogen	Temperature	Category	Hydrogen
Hydrogen	Temperature	Lower	0	-60		0.00%
0 – 100	-60 - 80	Limit 1	100	80	1	84.89%
100 – 700	80 - 100	Limit 2	700	100	2	15.11%
700 – 1800	100 - 120	Limit 3	1800	120	3	0.00%
1800 – 5000	120 - 140	Limit 4	5000	140	4	0.00%
5000 – 10000	140 - 250	Upper[14]	10,000	250	5	0.00%

Table 5-4 shows the likely membership of each category for the two measurement values in Table 5-3. Both readings are near a category boundary, but the error in the temperature is stated to be half of that for the hydrogen reading. Consequently, it has a much tighter distribution and less membership across the category boundary, i.e., ~2% as compared to ~15%.

One pair of measurements (hydrogen and temperature) and their estimated distribution allows for candidate values to be generated for collation into a final analysis of AHI. A Monte Carlo simulation could generate another 1000 pairs of hydrogen and temperature values by using the two actual measurements, the other statistical information from Table 5-3, and assuming a Normal Distribution.

[14] In theory, the upper limit is not defined

For each pair of generated measurements, the category into which the measurement falls can be identified. For example, if the randomly generated pair is (98, 102) as per Table 5-4, then the category pair will be (1, 3).

There are different options for creating a total score by combining the two categories for each, including:

- Evenly weighting the category scores, which is the equivalent of taking an average;
- Taking the maximum category as that is the most urgent;
- Adding the scores together with weighting; and
- Summing the scores in some other way.

There are multiple options to display a final figure: as a category, as a score out of 5, as a percent, etc.

The 1000 pairs of scores that were generated from the single pair of measured results and the associated statistical information creates 1000 total scores. These 1000 total scores form a distribution, giving an indication of the uncertainty of the result.

5.6.2 Linear Weighting – Parameters Each Contribute 50%

Figure 5-8 shows how the scores distribute for the evenly weighted average.

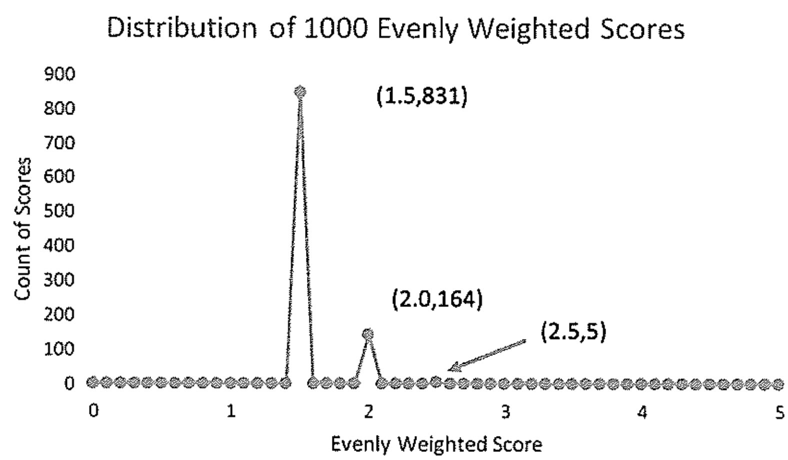


Figure 5-8: Distribution of Two Parameters, Evenly Weighted

The result of the analysis is predominantly a final score of 1.5, with some cases yielding a 2 and a few (0.5%) a score of 2.5. The variability of results reflects the uncertainty/error in the measurements and is a natural part of the measurement process. The categories are precisely defined, and so the pair of measurements could be in one of two or three categories, when the uncertainty of the measurement is considered.

If the resultant score derived from using this system is 1.5, what is known about the original data? Based on the categories, there must be one parameter score at 1 and one parameter score at 2. As explained previously, each category should have a well-defined meaning that is related to probability and a timescale. Knowing that at least one of the parameters has been allocated to category 2 provides an indication of the appropriate timescale for action.

Similarly, if the final average weighted score is a 2, the original parameters could have been categorized as follows:

- Both parameters could have been allocated to category 2, or
- One parameter could have been allocated to category 3 and the other to category 1.

Unfortunately, based only on the total score, it is unknown if there is actually a parameter in category 3, meaning that the timescale or urgency for action is also unknown. If the raw data was accessible, the categories could be checked and an urgency assigned to the final score.

5.6.3 Maximum Category Approach

Another approach is to look at the maximum score in each measurement pair. Figure 5-9 shows the distribution of 1000 measurement pairs, which are categorized using the scheme in the previous chart, and then the maximum score is used to give the final overall category.

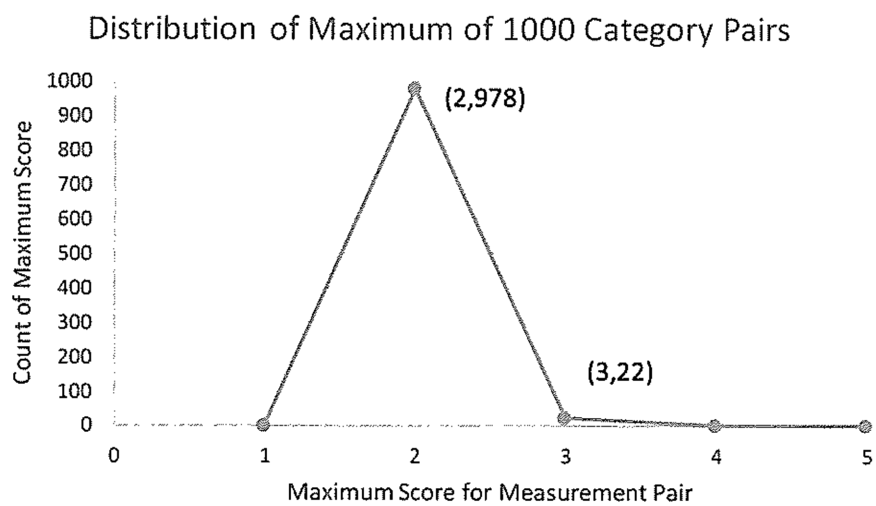


Figure 5-9: Distribution of Maximum Score of 1000 Category Pair Scores

Almost 98% of results are a category 2 and about 2% are a category 3. As long as there is an estimated PoF or range of PoF for an individual category, then this approach yields the parameter that is categorized as having the highest PoF and therefore requires the most urgent attention and level of response. It would be necessary to access the raw data to determine which parameter/s were in category 2 and determine what types of responses might be required in the appropriate timescale.

The maximum category is, for the measurement pair data, generally higher than the result from the weighting approach. A common effect of weighting or summing scores is to dilute or average out any effect of individual categories, potentially masking problems that need attention more urgently.

Choosing a method to combine two or more categories is mostly a matter of preference and, as per the introduction to Section 1.0, the choice should reflect the ability of the final AHI to address the initial question.

5.6.4 Effect of Different Weights for Categories

Can reweighting categories change the outcome of the weighted analysis? In the Monte Carlo Analysis of the measurement pairs, setting the hydrogen weighting at 80% and thus the temperature weighting at 20%, a rerun of the analysis yields the data in Figure 5-10.

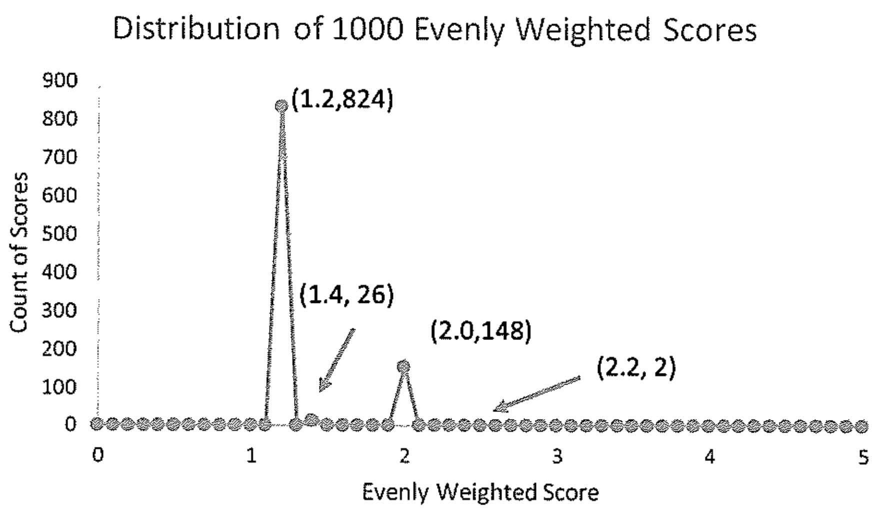


Figure 5-10: Weighting Hydrogen at 80% of Final Score

The reweighting has not affected the overall distribution greatly. There are different opportunities for scores, as the weighting multipliers are now dissimilar. Regardless, the same criticisms still apply:

- What does a score of 2 mean?
- What were the original categories?
- How urgent is the situation?

The Maximum Category approach has hardly changed. There is some variation in the number of generated measurement pairs with each of the possible maximum scores, as shown in Figure 5-11. Note that the Maximum Category is independent of weightings.

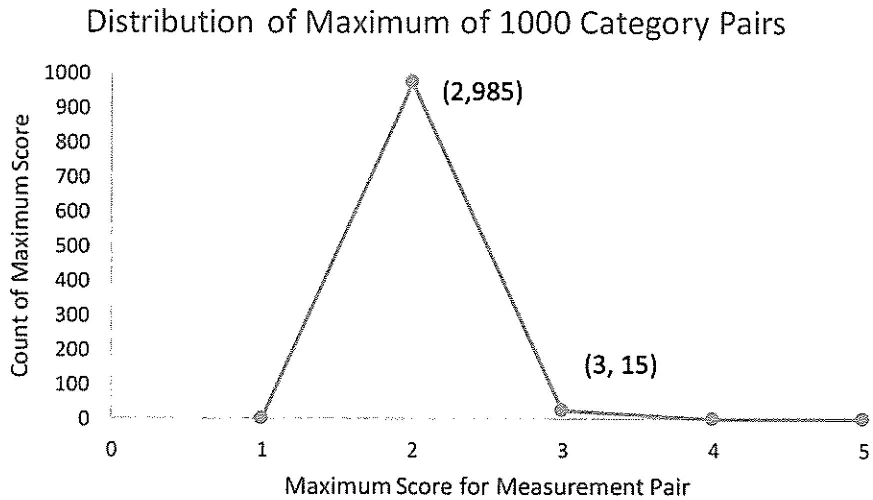


Figure 5-11: Maximum Category Distribution with Reweighted Categories

The Maximum Category is independent of weightings since it is calculated before weights are applied.

5.6.5 Using Log-scales for Category Labels

The aim of a log-based system is to maintain an indication of urgency.

For two parameters, a log scale approach can also be used. Table 5-5 shows a base 10 approach and an approximation of a base 3 approach. Instead of weighting individual measurements, scores are summed.

Table 5-5: Category Calibration

Linear Category	Log Category Base 3	Log Category Base 10
1	1	1
2	3	10
3	10	100
4	30	1000
5	100	10000

In the Base 3 system, any score over 100 is worthy of immediate attention; in a Base 10 system, any score over 10,000. That is, in both cases, any score that involves a measurement in Category 5 is worthy of immediate attention. In a Base 3 system, three scores of one category add up to be equivalent (in some cases approximately) to the next category up; in a Base 10 system, 10 lower categories add to the next higher category. Base 10 systems are also generally enumeration systems. Note that with these logarithmic systems, particularly the base 10 system, it is difficult for the total score to hide or mask the effect of a bad individual score.

A Base 3 Monte Carlo simulation of 1000 measurement pairs with no weighting yields a distribution of summed scores as shown in Figure 5-12.

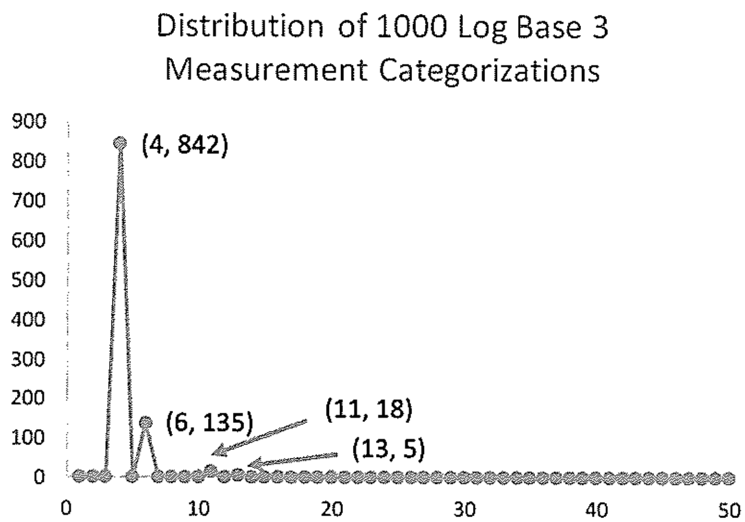


Figure 5-12: Log Base 3 Summed Score Distribution

The same issue of score interpretation is still prevalent. What does a score of 13 or 11 mean? There is, however, a clear indication of urgency, as discussed previously.

A Base 10 approach has a similar effect to a Base 3 approach but, if the contributing categories are limited in number (to 9 or less), it has a useful enumeration effect. The final score gives the number of contributing categories at each level. As an example, rather than use 2 parameters, let us take 8 parameters with base 10 scores as shown in Table 5-6:

Table 5-6: Base 10 Example—8 Parameters Evaluated

Parameter	A	B	C	D	E	F	G	H
Category	10	10	1	10000	100	100	1000	10

The final score, obtained by summing the individual categories, is 11231, reflecting the Base 10 approach. The benefit of the scheme is that each digit of the score represents a category. There are 2 contributing individual scores with a score of 100 (Category 3), 3 with scores of 10 (Category 2), and 1 each of 10,000 (Category 5), 1,000 (Category 4), and 1 (Category 1). Each score is both a reflection of the urgency and an encoding of the raw data. A score of 00161 has a highest contributing category score of 100 for one measurement, and has 6 other scores above the most benign category.

5.6.6 Urgency, Timescale, and PoF

The enumeration approach has been implemented by using timescales that are associated with each category for action. The timescales do not give a PoF directly, but they do imply urgency, and the associated timescale can indicate an implied PoF. If level 10,000 has an action timescale of 1 week, that might imply a likelihood of failure 100 times greater than normal. This cannot be calculated in any way: it is only an estimate based on knowledge and industry experience.

A probability of failure can be nominated or defined for each category, based on industry experience. When combining categories from different measurements, the key factor is that each category, for any measurement, should have the same associated timescale. This allows for planning actions, such as intervention with maintenance, refurbishment, or replacement, to be planned in a coordinated manner.

The implementation of the Log 10 enumeration scheme employed an asterisk, such that the system was designed for long term capital planning but any score which indicated an urgent review would have an asterisk on it, allowing for a short-term tactical response in a strategic framework.

The asterisk approach can be used with any system to highlight those assets that need urgent attention.

5.6.7 Extending the Experiments: 4 Parameter Monte Carlo Simulation

The two parameter Monte Carlo simulation can be extended to use 4 parameters, as shown in Table 5-7, with weightings toward a final total.

Table 5-7: Four Parameter Measurement and Distribution Statistics

	Hydrogen	Temperature	PD (1-100)	C2H2
Reading	105.00	95.00	46.00	4.00
Error +/- %	10	5	10	10
Lower	90%	90%	90%	90%
Upper	94.50	90.25	41.40	3.60
Confidence	115.50	99.75	50.60	4.40
# std dev	1.64	1.64	1.64	1.64
Std Dev	6.38	2.89	2.80	0.24
Weights	20%	30%	20%	30.0%

The parameter limits for categories are given in Table 5-8.

Table 5-8: Four Parameter Category Limits

Parameter	Hydrogen	Temperature	PD	C2H2	Category
Lower	0	-60	0	0	
Limit 1	100	80	50	2	1
Limit 2	700	100	75	5	2
Limit 3	1800	120	90	10	3
Limit 4	5000	140	95	25	4
Upper	10,000	250	100	50	5

The resulting distribution of scores, as more parameters are added, begins to smooth out the distribution chart, as seen in Figure 5-13.

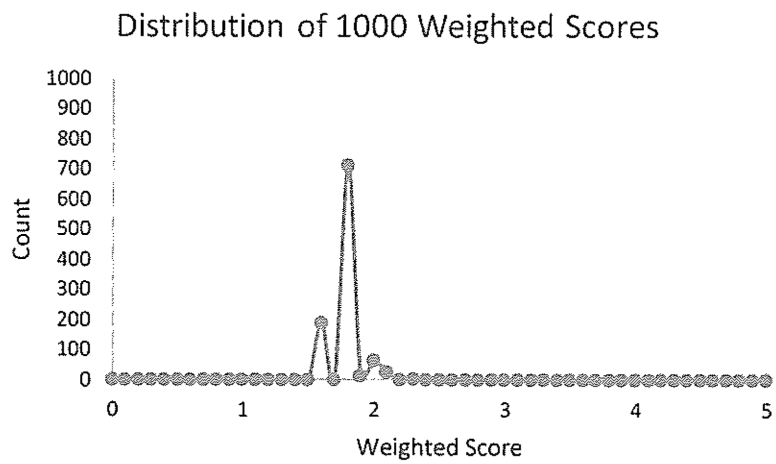


Figure 5-13: Four Parameter Weighted Summation Distribution

As more parameters are added, there are more uncertainties and more ways to generate an individual score using a weighted system.

Similarly, when a score changes, there are more ways that the change in score could have occurred. Clarity is lost as to whether there was a small change in a heavily weighted factor or a larger change in a less heavily weighted factor. The weighting system tends to dilute and average out the causes of variation.

The maximum category approach continues to summarize the maximum value in each quartet of measurements, as shown in Figure 5-14.

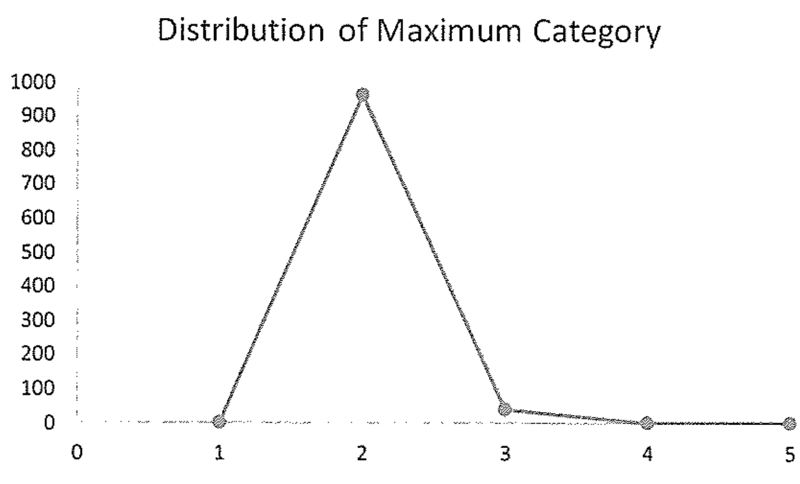


Figure 5-14: Four Parameter Maximum Category Distribution

A logarithm approach could help identify sudden changes of state for the quartets of results. A maximum value approach is likely sufficient to answer the question, assuming that there was a clearly defined question at the start.

5.6.8 The Final Score – What Was the Question?

The AHI should be set up to answer a question. The question and the precision of the desired response should be decided prior to developing the AHI.

After the AHI has been developed, check that the question is being answered appropriately. There should be no surprises, as long as the data and the analyses are understood.

5.7 Categories, Assessments, and Combinations

In general, aspects of an asset that are assessed as part of an AHI evaluation need to be classified in terms of whether they require tactical or strategic intervention and how they relate to the purpose of an AHI.

For example, tires may be included in a car health assessment, but the intervention to address tire issues is not the same as to address the viability of the car. An AHI that includes the tires may be skewed to address issues that are maintenance issues rather than longer-term replacement planning. The tires may be considered a subcomponent of the asset that need their own AHI. It is understood that if the tires fail, they put the viability of the whole asset in jeopardy. The AHI for tires might more reasonably be called an Asset Maintenance Index, or AMI. Note, however, that routine maintenance tasks, such as adjusting the tire pressure, should be done routinely in accordance with a well-defined maintenance plan. Routine maintenance should not be reliant upon a poor AHI result. The information from this routine maintenance task (e.g., one tire needs significant amounts of air added each week) can often feed into the AMI which will be used to determine if corrective maintenance, such as replacement of the valve, is required.

For a large asset, there may be physical subcomponents that are analogous to tires. For a transformer, these would include bushings, cooling system and radiators, On Load Tap Changer (OLTC), oil preservation system, etc. Each may be deserving of its own AHI, and may contribute to the overall AHI. It would be understood that the response to a bushing issue may be replacement, and for an OLTC the response may be corrective maintenance. Neither of these responses would contribute to the overall AHI for the main transformer active part.

Subcomponents may be physical devices, such as bushings, but they may also be logical groupings, such as dielectric capability or thermal performance evaluation.

The decision as to what is included in an overall AHI, and how it is included, is up to the designer of the AHI system, the question to be answered, and the approach to the question.

5.7.1 Sub-component Assessment Analysis

As an example of sub-components and the parameters being used to address them, Figure 5-15 shows 7 measurements that have been coded on a Base 3 logarithmic scale to yield a category, in green on the left. The example shows the possibilities.

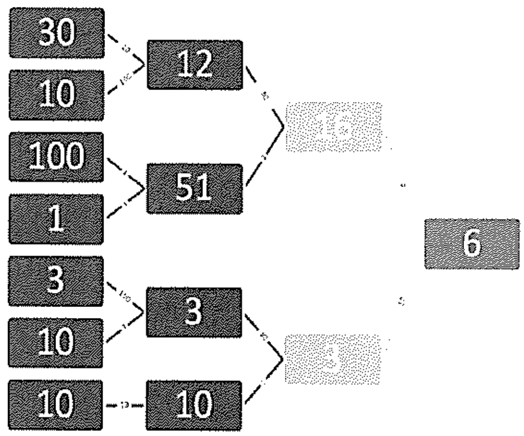


Figure 5-15: Multiple Sub-Component Analysis

Each measurement contributes through a weighted system to four sub-component assessments, in purple, which are normalized. Those sub-component assessments are then brought together into two super-component scores, which are again weighted and normalized. Finally, an AHI is generated.

Each measurement could contribute to several sub-components. Likewise, each sub-component could contribute to several super-components.

The level of uncertainty associated with each measurement contributes to the uncertainty at each subsequent level. The value of the AHI at each level answers a question appropriate to that level.

Although the original parameters were assigned to well-defined categories, each of which related to a timescale for any required action, it is unlikely that the final score, or the sub-component scores, can be easily and accurately related to a level of urgency or to a PoF.

5.7.2 AHI Effectiveness: A Cautionary Tale

This example is used to show that an AHI may seem to be providing value when, in fact, it is no better than random replacement.

As a simple example, 20 transformers are given scores for DGA, Winding Condition, and Age. Each criteria is scored on a scale of 1-5, and a weighted sum calculated (see Table 5-9).

Table 5-9: AHI Scores for 20 Transformers

I.D.	Age	DGA	Windings	Age	Weighted Sum
S1	49	3	1	4	3
S2	13	1	1	1	1
S3	48	2	1	4	2.9
S4	36	1	2	3	2.5
S5	41	5	1	3	2.6
S6	59	4	1	4	3.1
S7	3	1	1	1	1
S8	26	1	2	2	1.9
S9	59	1	5	4	4
S10	34	1	1	3	2.2
S11	32	1	2	3	2.5
S12	8	2	1	1	1.1
S13	34	1	1	3	2.2
S14	9	1	1	1	1
S15	61	3	1	5	3.6
S16	59	5	1	4	3.2
S17	20	1	1	2	1.6
S18	51	2	1	4	2.9
S19	21	1	1	2	1.6
S20	64	1	2	5	3.7
Average	36.35	1.9	1.4	2.95	2.38

The weightings for each measurement are 10% for age, 30% for windings, and 60% for DGA. The average of each contributing score is given at the bottom of the columns.

The overall population has an average AHI of 2.38.

What happens when some transformers are replaced at random?

The new transformer should be in pristine condition, so the age will drop to the minimum, and the DGA and winding scores will reset to new, low values.

In Table 5-10, two transformers have been randomly chosen and replaced, with the original weighted score and the new weighted score shown.

Table 5-10: Replacing Transformers at Random

I.D.	Weighted Sum	Replace?	Age	New weighted
S1	3	*	0	1
S2	1		1	1
S3	2.9		4	2.9
S4	2.5		3	2.5
S5	2.6		3	2.6
S6	3.1		4	3.1
S7	1		1	1
S8	1.9		2	1.9
S9	4		4	4
S10	2.2		3	2.2
S11	2.5		3	2.5
S12	1.1		1	1.1
S13	2.2		3	2.2
S14	1		1	1
S15	3.6		5	3.6
S16	3.2	*	0	1
S17	1.6		2	1.6
S18	2.9		4	2.9
S19	1.6		2	1.6
S20	3.7		5	3.7
Average	2.38		2.55	2.17

As seen in the table, the overall average condition of the population improves. This happens when weighted systems have age contribute to the health score.

The problem is that if an age-related system, as shown, is employed, it will self-justify. No matter what replacements are made, the overall population's health will improve, thus justifying the replacement.

To be both useful and of value, an AHI must do better than a system which replaces at random.

5.8 Generating an Asset PoF Derived from AHI

How can a PoF be generated from an AHI? Consider:

- The possible lack of clarity in the design of an AHI from the outset,
- The uncertainty in measurements made,
- The scarcity of definition in the functional relationship between parameters measured and PoF, and
- The averaging and dilution effects in weighted approaches.

Can the single AHI number be translated to yield the information needed? This may be possible using statistical analysis and justification of functions to transform an AHI to an Asset PoF.

It depends on whether the data is available and whether the AHI can be unpicked to regenerate the information needed. The AHI process removes and reduces information to clarify and simplify. The details that were taken out to generate the AHI in the first place need to be put back. The chances of recovering the details are better if the AHI is well-designed. There is little chance of recovering the needed details from a poorly designed AHI that provides a meaningless result.

There are two distinct approaches:

- A functional analysis that involves working backwards from an AHI via the combinatory functions to the raw data and the associated PoF; and
- A distributional approach, applying population statistics to AHI categories.

5.8.1 Functional Analysis

A functional analysis must be aware of how the AHI was set up, including:

- How were parameters codified?
- What timescales, if any, associate with the codes?
- What weightings or collation system has been used to produce a single number?
- What are the uncertainties on the codes, based on the codification?
- Has any PoF been included in the development of the AHI (i.e., in the codes)?

There are several steps, as shown in Figure 5-16.

Parameter	**Analysis**	**Coding**	**Subcomponent**	**Collation**	**AHI**
Measured with error and uncertainty	Levels, pattern match, delta, rate	Interpret on a scale: 1-5, say – with timescale	Collate data that will give a logical subcomponent	Generate an initial collated score	To a useful level or range, e.g. %

Figure 5-16: Identifying the AHI Steps

The steps need to be defined and laid out.

For example, a simple weighted system for four parameters, weighted such that one is twice as important as the other three and normalized to a percent, may result in something like Table 5-11:

Table 5-11: Simple Scoring System

Parameter	Analyses	Code	Weighting	Collated	Normalized	Delta
A	Level	4	20%			4%
B	Rate of change	1	40%	1.80	36%	8%
C	Level	2	20%			4%
D	Delta	1	20%			4%

The following information must be known:

- Accuracy of the initial measurements
- Accuracy of the analysis to give a code
- Meaning of the code either in a timescale or a PoF
- Weighting values—as shown
- Means for collation—in this case a cross product of the codes and weights
- Normalization process—in this case it is division by the maximum possible score

How to work back from a score of 36% to an AHI? One approach is to note that the minimum score possible, with all codes set to 1, is a score of 20%. What has caused the change from the minimum of 20% to 36%?

The contribution of a step change in each parameter is given as the Delta. A change of parameter C from a 2 to a 3 would increase the Normalized score by 4%. Via some math analysis, identify which parameters changed and by how much. Then, if the parameter codes are linked to a PoF, collate those PoFs to an AHI PoF.

The key is to know what timescales for action and what PoFs were applied to each code.

Assuming that the normalized PoF is a representation of general deterioration could be misleading. If a higher normalized percentage score has a lower set of codes, it should have a lower probability of failure, despite the higher normalized score. This is the result of non-linear weighting systems.

A MAX system could retain the coding information on timescale and PoF.

A Logarithmic system could retain the coding information on timescale and PoF.

Therefore, it may not be possible to relate an AHI to a PoF in anything other than broad general terms. It depends on how the AHI was designed and constructed.

5.8.2 A Distributional Approach

In a distributional approach, the assets are ranked as a population, or compared to another, representative group where failure statistics are known; those statistics are then used as a guide to PoF.

If the scores are normalized as a percent, what proportion of the assets should have a final score above 60%?

To apply the distributional approach, the AHI needs to have a meaning: what is the probability of failure for an asset with a normalized score above 80%? Above 70%? To derive these PoFs, the user could examine the scores from a year ago and relate the number of assets in each range at that time to the number of assets that failed. This at least gives some justification, based on history, to PoFs applied.

For example, if there is a historical failure rate of 1% per year for an asset class and 5% of assets have a normalized AHI above 80%, *at least* 1% of those would be expected to fail. Since these are the poorest performers, it is suggested that they should have a failure rate much higher than 1%.

Can this group be given a number? What proportion of those assets above 80% failed last year? In the previous year? What is the historical guide?

These are good points to discuss, but may be difficult to answer with any accuracy or precision.

The next section goes into the development of an AHI that can lead to a PoF with some confidence.

6.0 SUGGESTIONS ON DEVELOPING AN ASSET HEALTH INDEX

This section details some methods for developing an AHI that preserve the timescale and urgency information that can indicate a PoF.

6.1 Introduction

First question:

What problem is the AHI trying to address?

Second question:

Does the design of the AHI include time? In other words, does it include a sensitivity to action?

Data can be taken from multiple sources, manipulated, and used to produce a number; however, if calibrated timescales for action are not included at the start, they cannot be used. The timescales can be based on experience, or relate to known failure rates or industry expectations—but there still needs to be a time element.

Note that measurements have an accuracy. When they are coded, they may lose that accuracy.

Start small, keep track, and grow. An AHI is a model: an estimate of the asset's health or condition. It is a figure of merit that allows the relative health of assets within a group to be ranked, with the ranks defined by the model. An old rule of model development is that it is easier to build on a small model that works than to try to fix a large model that does not work. Consider that the business purpose of creating an AHI requires that it is easy to model and scale; this requires the model be constrained in complexity or risk not being usable, except by a very small group of consultants or aficionados.

6.2 Single Parameter: Good and Bad

This section will examine some aspects of an AHI system based on a single parameter: the measurement of hydrogen in transformer oil.

Two condition categories for transformers based on the parameter level are defined: Good and Bad. With each condition code there is a well-defined, if possibly unrealistic, action and timescale for action. The analytic is based on a preset level. Exceed the level and the transformer has gone from Good to Bad, and is illustrated in Table 6-1.

In reality, there should be a lot more information before deciding to replace a transformer.

Table 6-1: Two-Category System: Good and Bad

Code	Hydrogen	Timescale	Action
Good	<100	1 year	Resample
Bad	>=100	1 month	Replace transformer

What is the PoF for each category? What is the assumed PoF? Maybe in a year 0.5% of Good assets fail. If there are Bad transformers that need to be addressed within a month, is the PoF within the month also 0.5%? If the risk is higher, we would be putting extra risk on the organization.

In this case, it is assumed that the PoF in the timescale categories are consistent, a rule of thumb which is understandable and somewhat based on experience and knowledge.

Note that PoFs do not scale simply. With one dice, the chance of throwing a six is 1/6. With six dice, the chance of throwing at least one 6 is not (6 * 1/6) = 1 = certainty, but (1-(1-1/6) ^6) or about 67.5%

Add some PoFs for the period under consideration, and calculate the annual equivalent, as shown in Table 6-2.

Table 6-2: Two-Category System: Adding PoF

Code	Hydrogen	Timescale	Action	PoF	Annual Equivalent
Good	<100	1 year	Resample	0.5%	0.5%
Bad	>=100	1 month	Replace	0.5%	5.8%

Note that the annual equivalent of a monthly 0.5% PoF is not 6%, but just below. It is possible to work back from a stated annual equivalent to the PoF in the period under consideration. If the value for annual equivalence is decided to be 3%, the annual PoF cannot be calculated as the data is not available – then the monthly PoF would be calculated as not 3/12 % (or 0.25% by direct division), but about 0.254% by appropriate math. 0.25% and 0.254% may seem approximately close, but these are for small numbers. Note the variation in the reverse calculation where 100% certainty reduced to ~67% when appropriate math was used.

The model can be extended by adding a Rate of Change analytic. Deriving a second parameter (and associated analytic) ***rate of change*** from the first parameter requires having the period for the change. How are the analytics calibrated? Calibration is the means by which there is a common base to compare different analytics. Each category needs to have the same implication for action for the same category value. All Good values should have a common timescale for action. What makes the hydrogen level <100 and the hydrogen rate of change <50% equivalent values in the table? There is uncertainty, but without the calibration on timescales for action, there is no chance of ***consistent response.*** Someone has to set the levels for which the hydrogen level goes from Good to Bad, and at which the hydrogen rate of change goes from Good to Bad, allowing for consistency in approach and urgency. Table 6-3 uses some values for illustration.

Table 6-3: Two-Category System: Two Parameters

Code	Hydrogen	H2 Rate	Timescale	Action	PoF	Annual Equiv.
Good	<100	<50%	1 year	Resample	0.5%	0.5%
Bad	>=100	>=50%	1 month	Replace	0.5%	5.8%

How should the two analytics be combined? Each can be Good or Bad, and it is possible that neither analytic is triggered, or that both are, or just one. How is a final condition code for the transformer determined? If this problem cannot be addressed with this small amount of data and two analytics, what chance is there of addressing it with a large data array? What happens when the analytics become more complex or fuzzy? If one analytic is designed to identify that an asset is in poor condition, then the presence of other analytics should not be allowed to dilute that and somehow ameliorate the situation. If there is only one analytic, the outcome would be simple: Good or Bad. If there are cases where the output of an analytic needs to be reviewed because of other analytics, then the analytic should be redesigned or the initial analytic replaced using combinatorial analytics. The analytic scoring systems do not need to be correct; they are needed to pragmatically achieve the business goals.

Take, for example, the worst case as the actual condition. The table could extend to 4 overall states, as per Table 6-4. If the only value that is known is hydrogen, a decision is clear. If there is a Bad hydrogen result, and a Good H2 rate is added, the health been improved or the original analytic confirmed? The results in Table 6-4 are based on the possible outcomes of the two analytics.

Table 6-4: Two-Category System: Four States

Analytic	Hydrogen	H2 Rate	Overall
Case 1	Good	Good	Good
Case 2	Good	Bad	Poor
Case 3	Bad	Good	Very Poor
Case 4	Bad	Bad	Bad

This table shows the *dilution effect* and a result of weighting systems (in this case a uniform linear weighting). Now there are two categories that were not covered by the original data: Poor and Very Poor. Adding in more factors could result in dozens, if not hundreds, of new categories, while initially the only desired categories were Good and Bad. Not only has the final analysis been smoothed out, but even more outcomes and perhaps timescales for action have been created.

Care must be taken here. There is an advantage to having labels such as Good and Bad. There is no ability to add and average scores, which is easy to do with numeric labels. How this situation is dealt with is crucial to the creation of scalable models that deal with large amounts of data, or more granularity to the coding analysis.

Now add a third level of coding: the "Ugly," which has an action timescale of 1 day. Consistency and equivalency will be maintained by assigning a PoF for that day of 0.5%, as shown in Table 6-5.

Table 6-5: Three Category System: Estimated PoF

Code	Hydrogen	H2 Rate	Timescale	Action	PoF in Timescale	Annual Equiv.
Good	<100	<50%	1 year	Resample	0.5%	0.5%
Bad	>=100 & <1000	>=50% & <100%	1 month	Test/Fix	0.5%	5.8%
Ugly	>=1000	>100%	1 day	Replace	0.5%	~84%

Note how the annual equivalent PoF has gotten very large, underlining the urgency. It is also worth noting that there is choice in the PoF in the timescale for each code. There is no requirement for the PoF in each code to be consistent and equivalent. There is a requirement for the PoF to be monotonic: the annual equivalent PoF must rise as the period for action shortens, or it will be impossible to link the two in a meaningful way.

The advantage of a consistent PoF for the timescales is that it is intuitive to be able to compare the timescales on a PoF basis and that the urgency is not mixed into multiple timescales.

Note: Should assets be expected to deteriorate by stage, as with a Markov Chain approach? Such an approach may apply to condition-based failures, but not to random failures. Will everything go from Good to Bad to Ugly? The short answer is no, which makes a true Markov model difficult to apply. Work on such models is undertaken separately.

6.3 Labels: Log/Lin

The table below uses a linear label for condition coding. It is possible to generate an overall condition by summing, averaging, or weighting the individual scores. As noted in Section 3.0, this would lead to a loss of meaning in the resulting number with respect to urgency. Table 6-6 gives three category codes and limits.

Table 6-6: Three-Category System: Linear Codes

Code	Hydrogen	H2 Rate	Timescale	Action	PoF in Timescale	Annual PoF Equivalent
1	<100	<50%	1 year	Resample	0.5%	0.5%
2	>=100 & <1000	>=50% & <100%	1 month	Test/Fix	0.5%	5.8%
3	>=1000	>100%	1 day	Replace	0.5%	~84%

For example, summing the two codes for a transformer will result in a number between 2 and 6.

If a transformer scores 2 for hydrogen and 2 for hydrogen rate, does it have the same urgency as one that is 3 for hydrogen and 1 for hydrogen rate? They have the same summed score: 4.

A linear-average of the scores results in two scores of 2. The problem is that the story of the transformer's condition is markedly different in these cases. A log-based scale, as per Section 3.0, allows more urgent scores to stand out.

Therefore, the table can be extended to a set of log scale scores between 1 and 100, with a base 3 approach, as shown in Table 6-7.

Table 6-7: Five-Category System: Log Codes

Code	Hydrogen	H2 Rate	Timescale	Action	PoF in Timescale	Annual PoF Equivalent
1	<50	<50%	1 year	Resample	0.5%	0.5%
3	>=50 & <150	>=50% & <100%	6 months	Review/Assess	0.5%	0.998%
10	>=150 & <500	>=100% & <150%	1 month	Offline Test	0.5%	5.73%
30	>=500 & <1000	>=150% & <200%	1 week	Maintain	0.5%	23%
100	>=1000	>=200%	1 day	? Replace	0.5%	~84%

The increased number of categories allows for more granularity in analysis. The consistent PoF per timescale is useful in understanding what the timescale means and the urgency of action. Note that the scale of the analytic analysis does not need to follow either a linear or a logarithmic scale per the selected scale: that would be an artificial scaling. What is important is that the timescale chosen correlates to the analytic limits, not the other way around, and is monotonic.

There is no formal link from actual transformer condition to PoF: the process involves heuristics, experience, standards, and guidelines. In addition, a means to collate the two analytics for a single parameter into an overall AHI should be developed.

For a set of transformers, how are individual scores collated in order to give a meaningful AHI and a meaningful PoF? The extra data from the second parameter may provide confirmation of the suspect condition. Should two Bad scores be worse than one Bad score? This would seem intuitive, but may not be true—the scores may just be different ways of looking at a single problem; there is an increased precision (or confidence) in the score.

For example, Table 6-8 looks at 5 transformers ranked with a max score and enumeration condition code. The enumeration score is a count of how many 100s there are, then how many 30s there are, etc., and is a means to prioritize. The score is essentially an Asset Health Index (AHI), and it has meaning as the number of components that make the score and the associated PoF are evident.

Table 6-8: Five-Transformer Evaluation

Trf #	Units w/Hydrogen Score	Units w/H2 Rate Score	Max Score	Timescale	Enum Score	PoF in Time-scale	Annual PoF Equivalent
A	1	1	1	1 year	00002	0.5%	0.5%
B	3	30	30	1 week	01010	0.5%	23%
C	3	10	10	1 month	00110	0.5%	5.73%
D	100	3	100	1 day	10010	0.5%	84%
E	100	100	100	1 day	20000	0.5%	84%

Is a transformer with two scores of 100 more likely to fail than one with a score of 100 and a score of 10? The data may allow for a more confident analysis and may improve the precision of the estimate. What if the data for the hydrogen rate is missing? Is the probability of failure less? Does adding the Bad result make failure more likely?

The basis of the log 3 scale is heuristic. Three of a lower level score (e.g., 3 scores of 10) are approximately equivalent to one score at the next level up.

The number of entries in the table above could extend to include hundreds of data points, each with their own analytic or analytics. Then, a maximum score and an enumeration for ranking could be generated. The enumeration will become unwieldy, as there would be too many entries to manage sensibly.

Using components and failure modes means the hundreds of data points can be divided into far more manageable chunks. A transformer AHI that is based on a small number of components would work well with a combined MAX and Enumeration approach. Grouping by failure modes also infuses the process of developing analytics, and subsequent AHI, with scalability. It also makes the AHI generation process more manageable and easier to explain to the wide range of subject matter experts that will be needed to cover a large breadth of data. Whatever method is used, it is important to retain the urgency and the inherent ranking of the root data.

6.4 Analysis of Components and Failure Modes

A way of dealing with multiple data sources is to address components individually. With transformers, those components could be bushings, the cooling system, a tap changer, etc. Some components can be addressed through maintenance while others may need programs for refurbishment or replacement. As it is possible to get very granular in this analysis, it must be approached with the user perspective in mind. Components should be chosen to provide visibility to problems without trying to dissect the failure mode of a radiator bolt.

See Section 5.7 for an analysis of components and failure modes.

Think about the question that is being addressed: What does the AHI number mean? If it is to identify maintenance activity as an intervention, the timescale will likely be much shorter than a long-term replacement program. As with the car analogy, the entire car does not need replacement when tire pressures start to reach low levels: the tires simply need to be maintained.

Analysis of components follows the same process as analysis of the whole asset: data, timescales, and PoF. The scores for assets can be used as the raw data for the overall asset. Table 6-9 indicates some of the possible subcomponents of a large power transformer asset.

Table 6-9: Transformer Components

Power Transformer Components	Note
Tank	May be maintainable to address corrosion
Oil	Oil may be maintainable; also contains diagnostic data on other components
Cooling/Radiators	Likely maintainable
Tap Changer	Likely maintainable
Deluge System	Maintainable
Main windings	Likely not maintainable
Solid Insulation	Likely not maintainable – depends on location
Cable boxes	Maintainable
Bushings	Replaceable, likely not maintainable
Etc...	Oil containment, controls, and other devices

Each component score feeds into an overall assessment of the transformer; this is a common approach. It adds a level of complexity in determining components, but can add clarity, allowing for discrimination between actions such as maintenance and replacement at the component level. This approach more closely mimics the day-to-day way the different parts of the transformer are referred, and is ultimately more manageable.

To maintain calibration within the overall AHI system, timescales in the component analyses must be consistent across different components, or else a score of 3 may mean different things for different components, which means they cannot be combined sensibly.

The application of calibration results in a means to collate data through analytics to assess components, then a means to collate component data into an overall AHI. By building time into the approach, the PoF can be indicated both at the component and at the asset level.

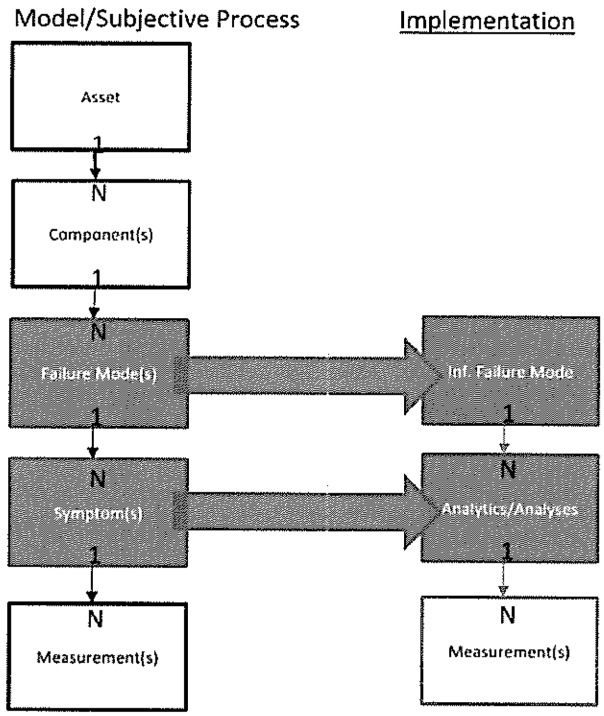
6.5 Failure Modes

If using a log scale, look for maximum scores and build in time for action based on acceptable base failure rates to indicate a PoF for the overall asset in a justifiable manner: a chain of reasoning should go from measurements through analytics to a PoF. The problem is that the approach is heuristic, and there is little corroborative information that a hydrogen level of 125, say, is really well correlated with PoF.

This can be addressed, to a degree, by using failure mode analysis.

Failure modes, which are generally well understood through RCM and similar analyses, can be linked to components and to the asset leading to the creation of an audit trail of sorts.

Figure 6-1 shows the basic theory and the subsequent implementation.



In summary for the model:

- An asset has 1 or more components
- A component has one or more failure modes
- Each failure mode has one or more symptoms
- Measurements can be made to confirm the presence of a symptom

In a practical implementation:

- We make measurements and collect data
- We analyze data to infer symptoms
- We analyze symptoms to infer the failure mode

Figure 6-1: Failure Mode Model

As an example, consider bronchitis as a failure mode for humans. The example shows how complicated things can become if one does not keep track of the details:

In summary:

- The human has many components, of which one is the respiratory system.
- The respiratory system has many failure modes, some of which are maintainable. Bronchitis is one such failure mode.
- There are many symptoms of bronchitis, one of which is the presence of a fever.
- Fever has many indicating measurements, one of which is oral temperature.

In practice, several measurements related to temperature may be interpreted, as illustrated in Figure 6-2.

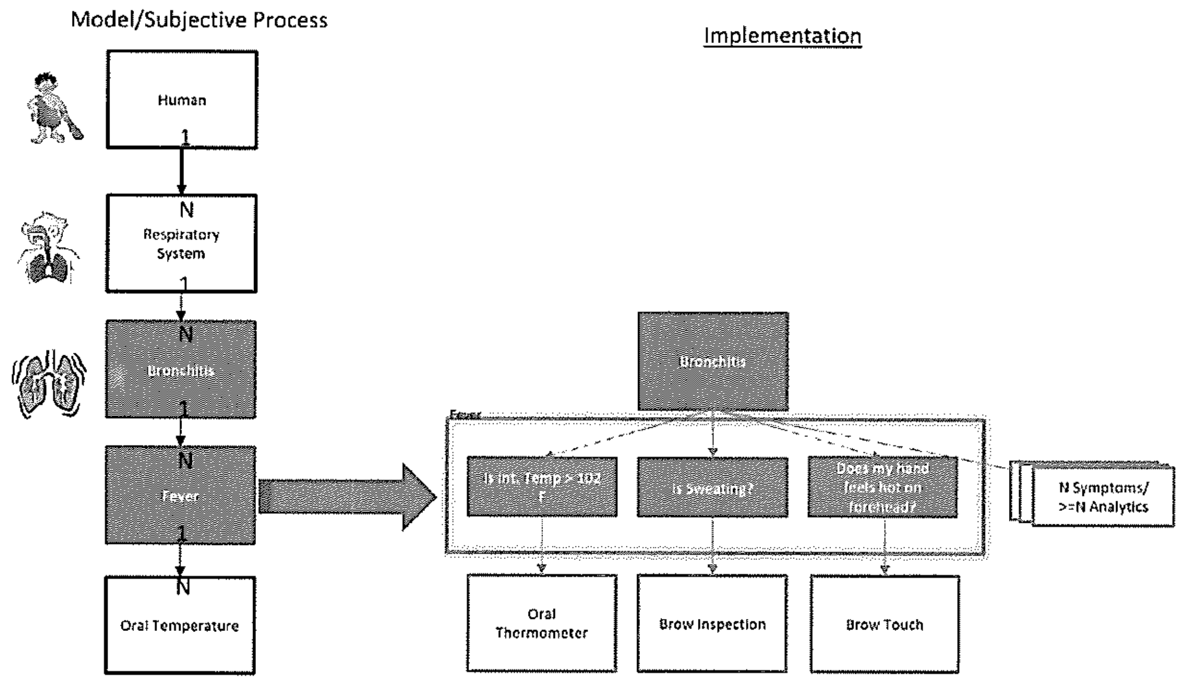


Figure 6-2: Failure Mode Model: Implementation

The figure shows the use of three data points and three analytics that can be used to infer the presence of a fever. For each data point, one or more analytics are associated—some analytics may use data for level comparison, or for ratio analyses or for more complex treatment.

When applied to an apparatus, the approach brings a richness to the measurement analyses and the subsequent AHI. The approach does not move from data through intuition to a PoF: it moves from symptoms (based on data) to failure modes to identifying issues with components. The components are then brought together in an overall AHI. That AHI may be used to address a reduced number of components—e.g., having a replacement index for the transformer, but a maintenance index for bushings and tap changers.

Figure 6-3 below indicates the basic approach:

- There are n data sources, for individual gases, furfurals, and moisture.
  - Each data source may contribute to several analytics.
- There are m analytics, ratios, rates of change, levels, etc.
  - Each analytic may contribute to several failure modes.
- There are p failure modes: PD, overheating, ageing paper, etc.
  - Each failure mode may contribute to several components.
- There are q components which contribute to the overall AHI.
  - Each component contributes to the AHI.

There is discussion in the industry about which components should be considered part of a power transformer when building an AHI, and which should be considered separate. National Grid UK, for example, treats the tap changer and bushings as separate, maintainable items that are dealt with in their own component strategies. The same approach could be applied to surge arresters and other apparatus.

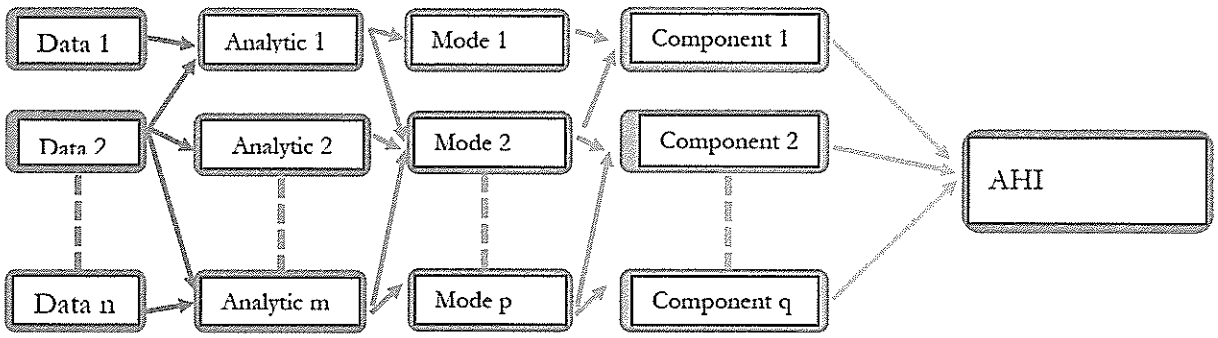


Figure 6-3: Failure Mode Model: Basic Approach

Failure modes are generally better understood, as they are real effects—in terms of what they are and the timescales for asset deterioration—as opposed to the basic symptoms of individual data points. Instead of saying "carbon monoxide is high, which is bad," say that "carbon monoxide is high and is a symptom of ageing which is the dominant failure mode for this asset and we expect it to deteriorate to a point requiring replacement over the next x years." One would never say that the transformer is failing due to a high carbon monoxide failure mode.

Analysis of transformer condition via components and failure modes follows.

6.6 Multiple Parameter Analyses and Identification of Failure Modes

The DGA analyses will be used as a means to discuss multiple parameter and analytic analyses, supporting failure mode identification, subsequent component assessments, and ultimately an AHI.

Transformer oil is analyzed to detect the presence and levels of several key gases, including hydrogen, acetylene, and others; there are standards to guide the interpretation of the gases and their implications [16].

The key gases can be used to determine the presence and severity of several failure modes, including partial discharge, arcing, overheating, and others. The diagnostics relate to transformer conditions rather than the diagnosis of the oil itself—the oil can be considered a component of the transformer in its own right, with its own failure modes and symptoms, and is considered a maintainable item.

The standards and guides available suggest analysis of dissolved gases by:

- Individual levels
- Rates of change
- Ratios and combinations of ratios (e.g., Rogers, as noted in [16])
- Ratios and combinations of ratios when all gases are above certain levels (Duval [16])
- % of combustibles (IEC Key Gas [16])
- Heuristic combined ratios—often *ad hoc* and purpose built systems
- Other approaches, including many R&D analyses

- Pattern Analysis (OLTCs)

Over several decades, there have been many analyses of available data which have shown that standards and guides can be misleading, and worse, inconsistent. Datamining approaches, which use neuro-fuzzy techniques and neural networks to track condition and deterioration of the power transformer over time [57], have been applied [56]. Standards and guides usually result in an indication of a particular problem: partial discharge, overheated paper, hot metals, etc.

The raw data for each dissolved gas are collated and the analytics are applied to:

- Identify level and rate of change codes, similar to Hydrogen in the earlier sections of this chapter; and
- Identify diagnostics from the application of industry standards and guidelines, such as Duval Triangles or IEC analyses, to indicate possible failure modes in operation.

So, the discussion of hydrogen would now be expanded:

- There can still be a level analytic which is coded 1 through 5 for each dissolved gas.
- There is still a rate of change analytic for each dissolved gas.
- Standard diagnostic analytics would be used, including Duval, IEC, Rogers, etc.
- Each diagnostic analytic would also have a 1-5 output for the diagnosed failure mode, with timescales calibrated to match the level and rate of change analytics.

With DGA, there is an array of analytics to detect anomalies and to diagnose particular fault conditions. Thus, prevalent failure modes can potentially be identified.

Illustrations of failure mode identification can be developed by considering DGA results. Consider 7 gases and some diagnostic analytics, as shown in Figure 6-4.

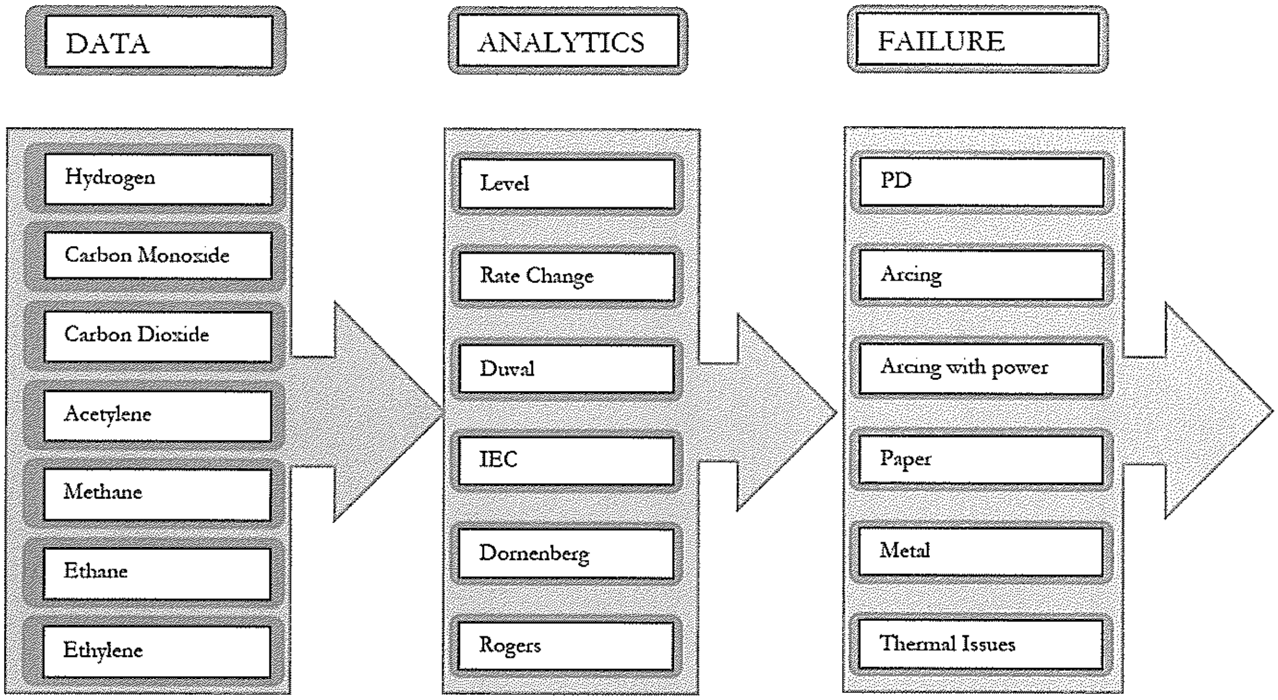


Figure 6-4: Some Transformer Elements: Data – Analytics – Failure Modes

Data – Analytics – Failure Modes

The figure shows the various data values on the left side, each of which feeds into the various analytics in the center; the list of analytics here is not exhaustive. The output of each analytic is a 1-5 value with a timescale calibrated to be consistent between the analytics. The actual diagnostics from Duval, Dornenberg, etc. do not usually include a timescale, but may have a severity indication. To generate the analytic timescale for each diagnosis requires the use of experience, industry heuristics, or some form of technical expertise.

The output of each analytic then inputs to the failure mode identification. The failure mode analysis is an analytic in its own right, taking in relevant information from all the available analytics. Individual analytics have a relevance to each failure mode analysis. This weighting is really based on experience: high hydrogen levels may be a significant and highly relevant indicator of PD, but a less relevant indicator of paper overheating. The relevance is used to confirm and support a diagnosis via the failure mode analytic. High acetylene is not usually a consideration for paper overheating; it has low relevance in that diagnostic.

The various failure modes can now be analyzed and collated, and thus the components that are in poor condition can be identified. Paper overheating is of low relevance to oil condition, being a more relevant indicator of solid insulation issues, while metal overheating is of low relevance to paper overheating. Failure modes relate to components based on a simple RCM analysis, one that does not require many levels of analysis but one that identifies the basic failure modes for each component. What does the component have which can fail?

With the failure modes in place, the AHI for each component can be identified by collating the failure mode outputs.

Scoring Process

To keep a consistent approach across each level:

- The output of each analytic is a time calibrated coding (1-5, 1-100, log, lin, alpha).
- The output of each failure mode analysis is a time calibrated coding.
- The output of each component analysis is a time calibrated coding.
- The final AHI is a time calibrated coding of each component score.

The consistent coding throughout makes the system easier to understand: wherever a 3 appears, the meaning is understood. Figure 6-5 summarizes this approach. As there is consistent scoring with consistent timescales related to consistent PoF, the PoF can be tracked at each level—however it may have been initially evaluated—and the PoF for the AHI can be inferred.

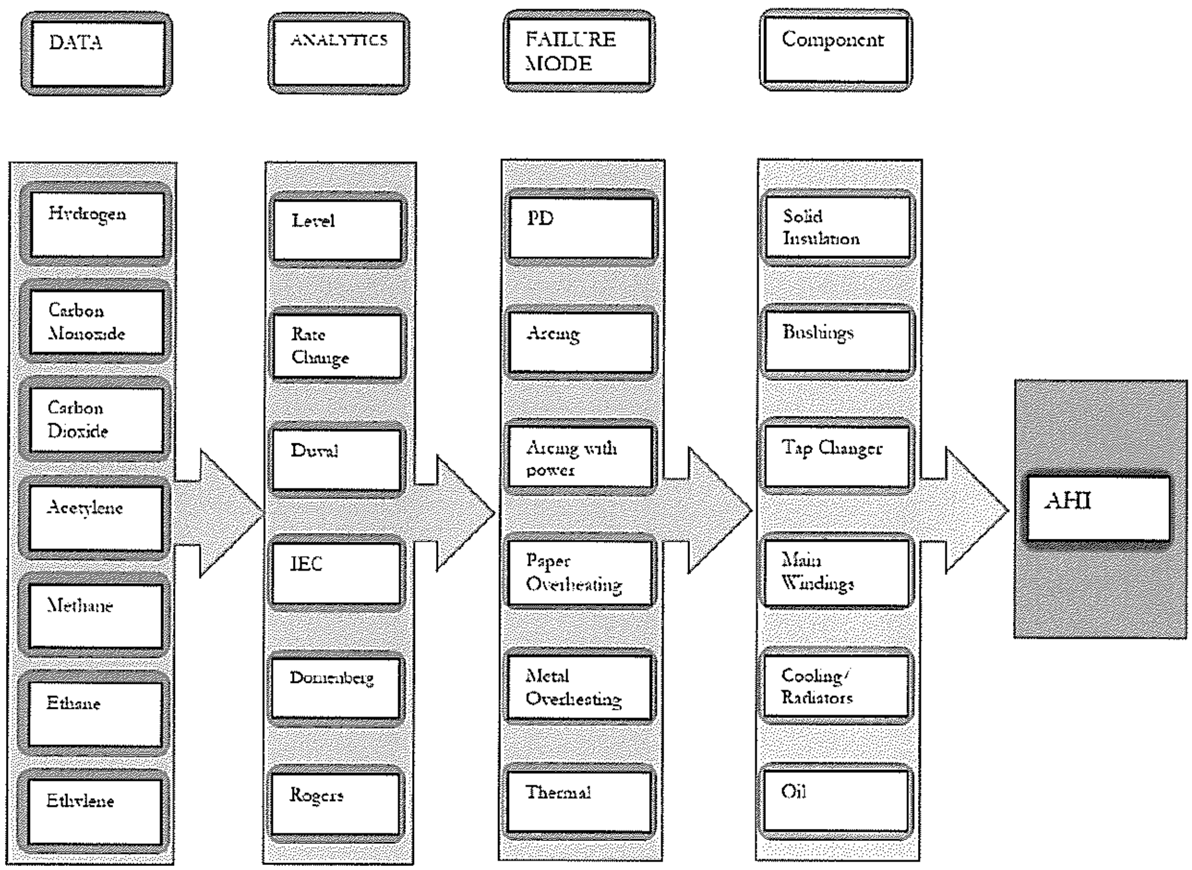


Figure 6-5: Some Transformer Elements: Extending to Components and AHI

6.7 The System in Action

There are elements to the process of generating an AHI with a meaningful PoF. Section 6.0 has covered some of the salient points and highlighted some of the causes for concern where uncertainty, inaccuracy, and imprecision are built in.

Steps:

1. Identify the assets of interest.
2. For each asset class identify the components (or asset subsystems, or whatever they are called in your system).
3. For each component identify high-level failure modes (a simple RCM analysis will suffice).
4. For available data, identify analytics (simple, standard, *ad hoc*) which indicate a failure mode in operation.
5. Score each analytic with a consistent and calibrated timescale of your choice, with each code/category labelled and assigned a PoF.
6. For each analytic, identify the relevance to each failure mode.
7. Collate analytics for each failure mode to score the failure mode action and timescale.
8. Collate failure modes for each component and score the component.
9. Collate failure modes for each component and score the asset.

Note that the collation of data must preserve the urgency of the analysis. The urgency of the analysis must be based on justifiable experience, statistics, and standards, within a risk management framework.

User Scaling of PoF and Codes

The system outlined here requires some idea of the base failure rate. For a large population, this may be a well-maintained statistic within the organization. For small populations, the variability may be significant and may require an estimate.

There may be a need to analyze the codes or categories and reassign an individual PoF for each category to override the time-based calculation. Figure 6-6 shows such an approach.

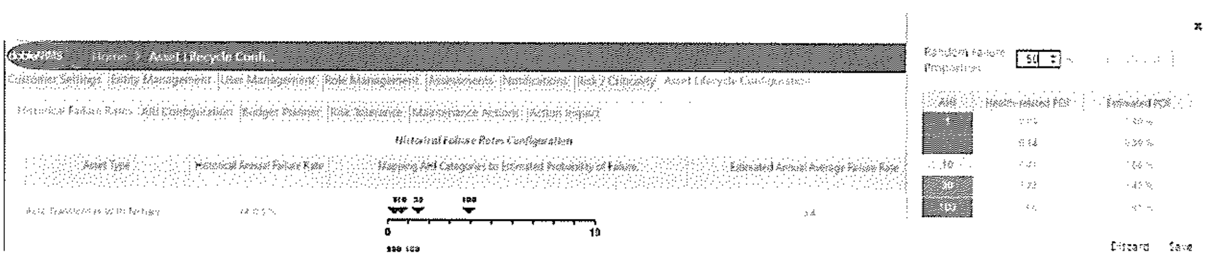


Figure 6-6: User Scaling of PoF

The historic annual failure rate is set and is editable—this is the target for the future.

Sliders allow the individual PoF for each code to be set: the resulting estimate is based on solving for a base case PoF and employing multipliers for all the other cases and numbers in each category. In Table 6-10, the sum of expected failures for the population total can be set to match historic failure rates. Thus X, the base case failure rate, can be calculated.

Table 6-10: Log Category PoF Back Calculation

Category	PoF	Number in Category	Expected Failures
1	X	N1	X * N1
3	3X	N3	3X * N3
10	10X	N10	10X * N10
30	30X	N30	30X * N30
100	100X	N100	100X * N100

It should always be possible for an AHI system to have direct user input to overrule estimated or calculated PoF values; however, it may become difficult to justify and relate all the manually entered values.

Addressing Risk Management

The system as described is not, at this point, a risk management system. There is a known PoF for the base case—e.g., the assets in code 1. That information can be used to address the PoF for assets in other categories and thus identify the number of assets that are likely to fail over and above the base case. If there was a consequence of failure—in monetary terms—the excess risk over and above the base case could then be calculated.

7.0 CONCLUSIONS AND RECOMMENDATIONS

There are many different ways to generate an AHI. Section 1.0 focused on motivations, constraints, and the use of AHIs, noting that there are multiple possible definitions for terms such as *failure* or *health*. In addition, the variability of data was noted, as well as the paucity of good links between measurable parameters and actual failures.

Section 2.0 provided a literature review, illustrating that there are many ways to develop an AHI and that these ways are not all similar or equally appropriate. The relationship between the input parameters and the output AHI can be tenuous, with many systems relying on an overall *average condition* approach, which used weights to combine individual parameter or component scores. Such approaches do not retain the timescales or urgency needed for action or for the possible derivation of a PoF, and can be very misleading if not downright erroneous. Assets do not generally fail through overall condition but through particular failure modes.

Section 3.0 reviewed the use of weighting systems in the development of an AHI and noted the drawbacks, including the difficulty of producing an AHI that is monotonic and that retains any sense of urgency based on the raw data analyses.

Section 4.0 reviewed practical systems that show the breadth of approaches which can be employed to address an AHI development. Many AHI systems do not have a clear statement of intent, and several AHI systems use complex math to seemingly hide the analysis under the guise of mathematical respectability. In some cases, there is a sleight of hand moving from an AHI directly to a PoF. The role of AHIs in justifying expense is common, so one may wonder at the ability of the regulator to dissect and critique some systems. Most AHI values reviewed in this report do not reflect the failure modes in operation or the likely interventions required. Furthermore, the urgency to address some failure modes is often hidden, possibly leading to an incorrect perception of the likelihood of failure.

Section 5.0 notes the difficulties of relating parameters to AHIs and finally to PoFs. The field is fraught with complexity and there are many opportunities to confuse and confound procedures; however, knowing the sources of variability allows for the development of an approach that may have some practical justification.

Section 6.0 goes through one possible approach to developing an AHI that can be used to create a justifiable and auditable PoF. The approach is somewhat heuristic, but is built on logic and can be tuned to reflect actual failure rates.

Overall, a good AHI will have an indication of the timescale for action and an audit trail that allows for justification of the actions to be taken. Ideally, an AHI will be built around failure mode analysis to provide a basis for action with calibrated timescales.

Many AHI systems use component scoring systems, which are often weighted, to produce an overall AHI. The use of log scales in such an approach helps preserve the timescales for action.

The use of age as an indicator of condition becomes a self-justifying approach, as shown by the fact that such systems usually improve the population's overall condition even if replacements are made at random.

What is needed is an AHI that reflects the failure modes likely to be present and relates those to timescales. This can be achieved through an analysis of known or expected failure rates. As asset populations may be small, the statistics of failures may be both imprecise and inaccurate; however, they are, at least, based in fact and can lead to a realistic ranking. Industry statistics, such as those available from CIGRE reports, are both useful and justifiable. It is also useful to disaggregate condition-based failures and random failures, or at least find a proportion of the overall failure rate that is expected to be condition-based. Given the weighted nature of many AHIs in use at present, it is suggested that few, if any, provide a basis for translation to a PoF.

The suggested approach, based on equivalent PoF for calibrated timescales, gives a rationale for a defensible analysis; in fact, the system can be tuned to reflect the needs of local conditions and experience. The meaning of the PoF for components and parameters, relating to timescales, is relatively easy to understand, and the preservation of timescales, to preserve urgency, allows for the derivation of a meaningful PoF.

Designing an AHI around timescales for action, and then building in the ability to estimate a PoF for each AHI category/code/value, is key to deriving a PoF from the AHI.

7.1 Working Forward to an Asset PoF

Working forward directly from raw data to derive an APF or Asset PoF is extremely difficult, if not impossible. The data required to determine the relationships is unavailable, the relationships are therefore uncertain, and the resultant outcomes are therefore imprecise. This may result in a ranking that monotonically reflects asset condition. Categorizing the population into a small number of condition bands, rather than having continuous AHIs which lead to more uncertainty in any derived PoF, may be a more informative way to describe the data.

7.2 Working Backward from an Asset PoF

In practice, failure rates for asset classes have been expected. If the population is assumed to look similar this year to what it did last year, roughly the same overall failure rates can be expected. The identification of the poorest performing units being associated with the worst asset health category/code is justified. The calibration of timescales can be preserved through equivalency of PoF for each category.

If an AHI system is built using the known or expected timescales for action, and the system is log-based, the estimated PoF can be preserved through the various steps to the final AHI. This is important in a business environment where the results are likely to influence long-term investment plans and where the analysis needs to be credible rather than act as a placebo.

It is not just the art of the practical that is needed, but also the art of the justifiable, auditable, and useful.

7.3 Opportunities for Improvement

A number of organizations may benefit from AHIs. However, expectations should be tempered by the availability of good failure data and by the need to work in an asset management and regulatory set of frameworks.

Given the wealth of knowledge on asset condition assessment that is available, an AHI should not contain any surprises. The authors thus recommend that any attempt at developing an AHI starts from what is well known and well understood, and that such an attempt grows from a simple start of a few parameters. Success can then be built in from the start.

8.0 REFERENCES

[1] McGrail, T., Heywood, R., " T. (2016). *Deriving a Useful Asset Health Index - Getting Started, Getting Value and Making Use of Them.* Doble Client Conference, 2016

[2] Hjartarson, T. (2003). Development of Health Indices for Asset Condition Assement. *IEEE T&DT&D Conference & Exposition, 2003 PES*

[3] CIGRE TB 642 "Transformer Reliability Survey", WG A2.37, December 2015

[4] Jagers J., Tenbohlen S., "Differences Approaches for the Acquisition of Reliability Statistics", CIGRE Paris 2009

[5] CIGRE TB 227, "Life Management Techniques for Power Transformer"

[6] CIGRE TB 660 "Saving Through Optimized Maintenance in Air Insulated Substations", WG B3.32, August 2016

[7] Ding H. & Ryder S., "Asset Health Review: How Good is it Really?", EuroDoble Colloquium, Manchester UK, 2012

[8] Publicly Available Specification (PAS) 55, Asset Management, BSI, 2008

[9] ISO 55000, "Asset Management", 2014

[10]*"Asset Management: An Anatomy"*, Institute of Asset Management, December 2015

[11]ISO 18095 TC108/SC5. (2015). Condition Monitoring and Diagnostic of Power Transformers. ISO

[12]McGrail, T., *"Asset Health Indices: What's the Point?"*, TJH2B TechCon, 2016

[13]CIO review on LinkedIn "Separating Fast and Risky in Analytics-Based Asset Management: a Utility Perspective", Vanessa Edmonds, VP, TMG Consulting, August 2016

[14] *"Subject Specific Guidance: Capital Investment, Operations and Maintenance Decision Making"*, IAM, August 2016

[15]Caussarieu A, Tiberghein A., *"When and Why Are the Values of Physical Quantities Expressed with Uncertainties? A Case Study of a Physics Undergraduate Laboratory Course"*, Intl. Journal of Sci. Math Educ., Springer, 2016

[16]IEEE C57.104 "The Interpretation of Dissolved Gas Analyses of Oil Filled Transformers", IEEE, 2011

[17]Blake, S. (2012). *Calculating a Health Index to Make Decision on Asset Replacement.* Strathclyde University , presented at Mathematical Methods Colloquium: Expert Judgment, Durham, UK

[18]DNO Common Network Asset Indices Methodology, UK Regulator, OFGEM, July 2015, Draft V3

[19]Feng Z., "Markov process applied to degradation modelling", University of Trondheim M.Sc. Thesis submission, 2013

[20]Hjartarson, T. (2006). *Condition Health Indices and Probability of Failure.* Kinectrics

[21]Hubbard, D., *The Failure of Risk Management: Why It's Broken and How to Fix It,* Wiley, 3rd Edition, 2014

[22]Hubbard, D., *How to Measure Anything,* Wiley, 1st Edition, 2009

[23]Tsimberg, Y. (2010). *Use of Asset Management Risk Assessment Methodology for Public Safety Analysis.* Kinectrics

[24]Zhou D, Wang. Z, Li C., 2013, *"Data Requisites for Transformer Statistical Lifetime Modelling – Part I: Aging-Related Features"*, IEEE Transactions on Power Delivery, Vol 28, No 3, July 2013

[25]Wang Y. et al, 2016 *Evaluating Equipment Reliability Function and Mean Residual Life with Continuous Monitoring,* IEEE PES

[26]Wang C., Hou Y., IEEE 2016, *A Novel Method for Optimal Life Cycle Management Scheme with Markov Model,* IEEE PES

[27]Coullon, J. L. (June 2014). Advanced Asset Health Managment for Effetive Maintenance and Asset Replacement Planning. *CIRED Workshop - ROME*

[28]Eric Dorison, F. L. (2007). Health Index. *Jicable*

[29]DiMatteo, S. (2015, 05 10). Big Data, Preditive Analtytics and Asset Helath Indexing. *Maint World*

[30]Edirisingh, R. F. (2009). Decision Making for Predictive Maintenace in Asset Information Managmenent. *Interdisciplinary Journal of Information, Knowledge, and Managmeent, 4*

[31]Albrice, D. (2015, April 24). *RDH.com.* Retrieved from RDH Making Building Better: http://rdh.com/are-my-assets-fading-or-degrading/

[32]Cantler, F. (2016). Vision 2050- An Asset Management Strategy. *IEEE PES*

[33]Transpower. (15 November 2013). *Asset Risk Management: Asset Health Framework.* Transpower New Zealand Limited

[34]Shepard, R.W. (2001, July-August). California Bridge Health Index, A Diagnostic Tool to Maximize Bridge Longevity Investment. *TR News,* 6

[35]Ryder S., Jarman P., Heywood R., "*Transformer Asset Management: How Well Are We Doing And Where Do We Need To Do Better?*", International Conference of Doble Clients, Boston 2015

[36]Exelon. (2015). *Franklin Park Distribution Asset Health Pilot*, October 21st, 2015

[37]SP Energy Networks. (March 2014). *SP Energy Networks 2014-2023 Business Plan Updated March 2014 Annex Asset Health and Criticality Strategy*

[38]Hydro One. (2005). *Hydro One Distribution - ACA Summary Report*. Hydro One

[39]Kydd T., *Aggregate Health Indices as Used for Asset Investment Decisions and Universal Understanding* EuroDoble, UK 2007

[40]Bennett, G. "*Transformer Condition Assessment*", Doble Client Conference 2003

[41]Kini M.V. *et al*, TATA Power Company Ltd., Mumbai, "Development of Transformer Health Index (THI) – TATA Power Experience", International Conference of Doble Clients, 2015

[42]Doble's "Life of A Transformer" Seminar, San Antonio, USA, 2015 - KeyNote

[43]Doble's "Life of A Transformer" Seminar, Miami, USA, 2014 – KeyNote

[44]UK Power Networks "*Asset Stewardship*" Annual Report, 2014, OFGEM, UK

[45]Kobayashi T., et al, *"Current Situation and Recent Challenges in Asset Management of Aging T&D Substation Facilities in Japan"*, CIGRE Paper B3_302_2016, Paris, August 2016

[46]Graves P. et al, "*Evolution of ComEd Asset Management*", T&D World My 2016

[47]Bosire E. & Yarrow A., Kenya Power and Lighting Co. *"Procedure for Using Condition Based Maintenance to Create the Health Indices of Transmission Power Lines: A Case Study of the Kenyan Coast"*, International Conference of Doble Clients, 2010

[48]Bosire E., *"Health Indices of Transmission Power Lines: A Case Study Mombasa Network"*, International Conference of Doble Clients, 2012

[49]Peck J., "*Leveraging Data for Long Term Returns*" Boston Water and Sewer Commission, Institute of Asset Management Boston Branch Meeting, September 2016

[50]Blanquez, F. *"Maintenance Strategy For The Oil Filled Transformers Supplying CERN's Particle Accelerators"* SEA Marconi Conference MyTransfo, Turin Italy, 2016

[51]Ali Naderian Jahromi, Ray Piercy *et al,* "An Approach to Power Transformer Asset Management Using Health Index", IEEE electrical Insulation Magazine, March/April 2009, Vol 25 No 2

[52]Stigler, S., "The Seven Pillars of Statistical Wisdom", Harvard University Press, 2016

[53] "Recent Developments in DGA Interpretation", CIGRE Technical Brochure 296, Pub. 2006

[54] "Mineral oil-impregnated electrical equipment in service - guide to the interpretation of dissolved and free gases", IEC 60599, 1999

[55]Nowlan and Heap, "Reliability Centered Maintenance", United Airlines Publication AD/A066 579, 1978

[56]McGrail A., *et al, "Datamining Techniques to Assess the Condition of HV Electrical Plant"*, Paper 15-107, CIGRE Session 2002, Paris, France

[57]Thang, K.F. *et al*, Application of Self-Organizing Map Algorithms for Analysis and Interpretation of Dissolved Gases in Power Transformers", IEEE PES Summer Meeting, 2001, Vancouver, BC, Canada, Vol 3, pp 1881-1886

[58]Bartley, W. "Analysis of Transformer Failures", 79th International Conference of Doble Clients, 2012, Boston, MA

APPENDIX A. Definitions of Failure

The information contained in this appendix is extracted entirely from CIGRE Technical Brochure 642, "Transformer Reliability Survey". It provides a reference to failure definitions from surveys, international standards, and guidelines from Annex A of the document, as well as a discussion about the most relevant failure definitions from CIGRE Technical Brochure 642, Section 3.3.1 titled "Failure."

The majority of the surveyed studies did not provide consistent failure definitions. In order to have a true comparison between the failure data, it would be required that the failure definitions be similar. This ensures that comparisons are based on failures (events) occurring under the same conditions.

The definitions of failure and associated terms and how these definitions are applied depends on the environment in which they were developed and are being used. The systems operator's focus would be on the impact on the system, ranking failure in terms of system reliability, whereas the plant specialist would rank it in terms of the remedial action required to restore equipment functionality. A clear example is that of transformers being removed in scheduled outages. The system operator would not consider this a failure, since it would not have an effect on the system reliability, whereas the plant specialist would consider it a failure. Another example is that of insurance companies that could be using insurance claims as a benchmark, where the definition of failure and its severity could be dictated by the value of the claim.

The definitions in IEC and IEEE are described as very broad, whereas Bossi's definition is more restrictive in that it considers only problems that require the unit to be removed from service for repair. Further examples of restrictive definitions include those of Kogan and Higgins. Despite the difference between a broader or restrictive definition, both types allow further breakdown of failures into levels of failure severity, or outage type and times.

ANNEX B: Definitions from Surveys and International Standards and Guidelines

SOURCE	TYPE	DEFINITION
[IEC, 1990]	Failure	The termination of the ability of an item to perform a required function.
[Bossi, 1983]	Failure	Lack of performance by a transformer of its required functions so that the unit must be taken out of service to be repaired.
[IEEE, 1986]	Failure	The termination of the ability of a transformer to perform its specific function.
[Cigré WG A2.18, 2003]	Failure	Any situation that requires the equipment to be removed from service for investigation.
[Kogan, 1988]	Failure	Any forced outage of a transformer due to its failure in service. Trouble which requires transformer to be returned to a factory for repair, or which requires extensive field repair. Transportation damage and minor troubles which may require an equipment outage are not considered as failures.
[Higgins, 2001]	Minor Failure	Can be repaired quickly in situ. The resulting outage typically would be less than one month.
[Higgins, 2001]	Major Failure	Must be repaired (if this is possible) off-site, usually at manufacturer's works. The resulting outage typically would be measured in months. Failures in which the transformer is destroyed or

		must be retired are also major failures.
[Bossi, 1983]	Failure with scheduled outage	Failure for which the transformer can deliberately be taken out of service at a selected time.
[Bossi, 1983]	Failure with forced outage	Failure for which the transformer must be taken out of service immediately (within 30 minutes).
[IEEE, 1986]	Failure with forced outage	Failure of a transformer that requires its immediate removal from service. This is accomplished automatically or as soon as switching operations can be performed.

APPENDIX B. Reference Material

from CIGRE Technical Brochure 227: "Life Management Techniques for Power Transformers"

This Appendix provides useful reference material from CIGRE Technical Brochure 227 (TB 227).

Failure Identification

The brochure contains recommendations on failure identification as shown below:

"Failure occurs when withstand strength of the transformer in terms of some parameter (dielectric strength, mechanical strength) is exceeded."

The situation is summarized in a figure.

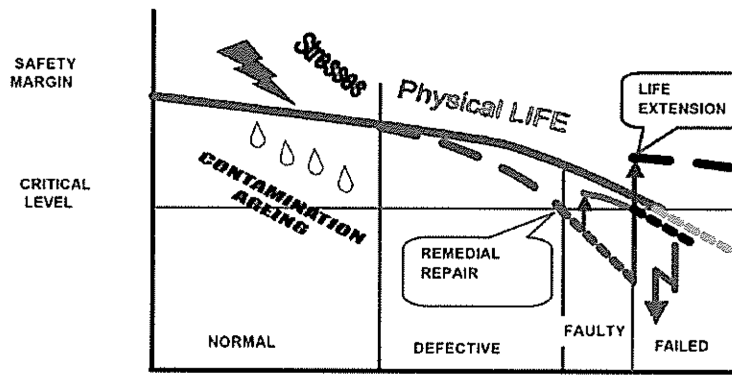


Figure B-1: Asset Deterioration (from CIGRE TB 227)

Failure Taxonomy

TB 227 considers that defects are reversible but faults are not reversible. A taxonomy is included in the brochure, as shown:

Table B-1: Failure Taxonomy (from CIGRE TB 227)

Code	Description
Normal	No obvious problems, no remedial action justified. No evidence of degradation.
Aged? Normal in service?	Acceptable, but does not imply defect-free.
Defective	No significant impact on short-term reliability, but asset life may be adversely affected in long term unless remedial action is carried out.
Faulty	Can remain in service, but short-term reliability likely to be reduced. May or may not be possible to improve condition by remedial action.
Failed	Cannot remain in service. Remedial action required before equipment can be returned to service (may not be cost effective, necessitating replacement).

TB 227 notes that after a failure, the failed item has a fault. Failure is an event, and a fault is the resultant state. TB 227 recommends referring to the state as a failed condition.

Defects are considered as non-conformance, whereas a fault is considered as deterioration beyond normal ageing/wear.

TB 227 also notes the IEC 60050 definition of failure: *"termination of the ability of an item to perform a required function."*

Condition Monitoring Strategy

TB 227 indicates that in its most general form, condition monitoring can be shown in the form of two loops, as illustrated by the diagram below:

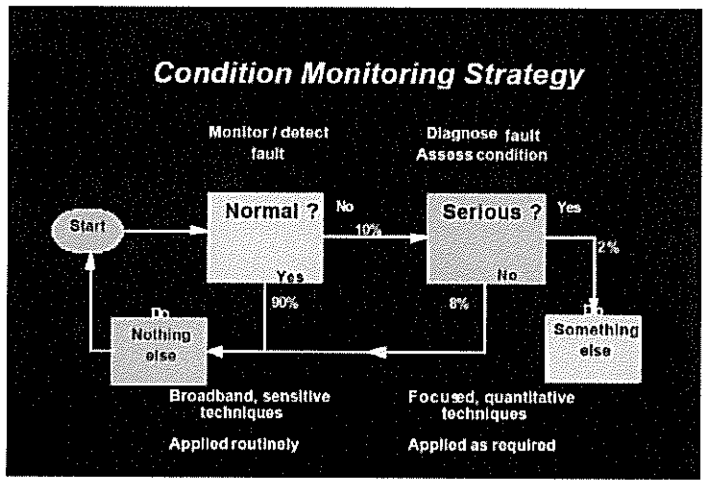


Figure B-2: Condition Monitoring Strategy (from CIGRE TB 227)

APPENDIX C. Extracts from TATA Power System

This appendix gives some details from the TATA power weighted AHI system.

Table C-1: TATA Weights for Condition Assessment

Fault gases	Wts.
Hydrogen (H_2)	2
Methane (CH_4), Ethane (C_2H_6)	4
Methane (CH_4), Ethane (C_2H_6), Ethylene (C_2H_4)	6
All above gases	8
All above gases plus Acetylene (C_2H_2)	10

Age (Years)	Wts.
0 to 10	4
10 to 20	8
20 to 30	12
30 to 40	16
> 40	20

Oil Acidity	Wts.
< 0.2	0
0.2 to 0.3	5
> 0.3	10

Electrical tests - IR / PI / Winding Tan Delta	Wts.
No Violation w.r.t iimits	0
Any one parameter Violation w.r.t limits	6
Any two parameters Violation w.r.t limits	8
Violation w.r.t limits	10

Furan content (PPM)	Wts.
< 2	0
2 to 3	8
3 to 4	12
4 to 5	16
> 5	20

Oil Resistivity & Tan Delta	Wts.
No Violation w.r.t Limits	0
Violation w.r.t Limits	10

Moisture content and BDV	Wts.
Moisture	10
Moisture and BDV	20

Table C-2: Scoring Weights for Transformer Components

Sr. No.	Transformer condition monitoring factor	Overall factor weightage (Wf)	Factor score (Sf)
C1	Dissolved Gas Analysis Factor (DGAF)	10	Factor rating derived from individual score & weightages of parameters / inspection check points as per Appendix-2
C2	Oil Quality Factor (OQF)	8	
C3	Furfural content Factor (FF)	6	
C4	Electrical Tests Factor (ETF)	10	
C5	Transformer Service Factor (TSF)	8	
C6	Inspection & Maintenance Factor (IMF)	6	

Table C-3: 4 TATA Weights for Condition Assessment

Parameter / Inspection check	Voltage level	Score (Si)					Wt. (wi)
		Good	Acceptable	Need attention	Poor	Very poor	
		4	3	2	1	0.01	
DGA factor							
Hydrogen (H_2)	All TRF	≤ 50	51-100	101-300	301-700	> 700	2
Methane (CH_4)	All TRF	≤ 55	56-110	111-200	201-600	> 600	3
Ethane (C_2H_6)	All TRF	≤ 30	31-65	65-100	101-150	> 150	3
Ethylene (C_2H_4)	All TRF	≤ 30	31-50	51-100	101-200	> 200	3
Acetylene (C2H2)	All TRF	0	1	2-35	36-80	> 80	5
Carbon Mono oxide (CO)	All TRF	≤ 200	201-350	351-900	901-1400	> 1400	1
Carbon dioxide (CO_2)	All TRF	≤ 1000	1001-2500	2501-5000	5001-7000	> 7000	1
TDCG	All TRF	≤ 365	366-676	677-1635	1641-3130	> 3130	2
Oil Test Factor							
Moisture content (in PPM)	U ≤ 145kV	≤ 20	21-40	40-45	46-50	> 50	4
	145 < U < 400 kV	≤ 15	16-20	21-25	26-30	> 30	
	U ≥ 400 kV	10	10-15	16-20	21-25	> 25	
Acidity of oil (in mg KOH/g)	All TRF	≤ 0.1	0.1 -0.2	0.2-0.25	0.25-0.3	> 0.3	3
Break Down Voltage (in KV)	U ≤ 145 kV	≥ 60	50-60	40-50	30-40	> 30	2
	145 < U < 400 kV	≥ 60	50-60	40-50	35-40	> 35	
	U ≥ 400 kV	≥ 60	50-60	45-50	40-45	> 40	
Dielectric Dissipation Factor at 90	All TRF	≤ 0.1	0.1-0.2	0.2-0.3	0.3-04	> 0.4	2
Specific resistance (resistivity) at 90 deg. C (MF : 10^12)	All TRFs	> 0.2	0.1 - 0.2	0.075-0.1	0.05-0.075	≤ 0.05	2
Furan Factor	**All TRF**	**≤ 0.5**	**0.5 – 2.0**	**2.0 – 3.5**	**3.5 – 5.0**	**≥ 5.0**	**6**
Electrical Testing							
Winding Tan Delta @ 20 Deg. C	All TRF	≤ 0.5	0.5-1.0	1.0-2.0	2.0-5.0	> 5.0	3
Winding Capacitance w.r.t. factory value / previous test results	All TRF	≤ ±5%	±5 - ±10%	±10 - ±15%	±15 - ±20%	> ±20%	2
Rate of rise of Winding Tan Delta @ 20 Deg. C w.r.t previous results	All TRF	< 5%	5-10%	10-20%	20-35%	35-50%	2
IR - HV winding (MΩ)	All TRF	> 2* KV	> KV + 1	> 0.8 (KV+1) to ≤ (KV+1)	> 0.6 (KV+1) to ≤ 0.8 (KV+1)	≤ 0.6 (KV+1)	1
IR – MV winding (MΩ)	All TRF						
IR – LV winding (MΩ)	All TRF						
PI - HV winding	All TRF	>2.0	1.3-2.0	1.3-1.0	1.0-0.5	<0.5	2
PI – MV winding	All TRF	>2.0	1.3-2.0	1.3-1.0	1.0-0.5	<0.5	
PI – LV winding	All TRF	>2.0	1.3-2.0	1.3-1.0	1.0-0.5	<0.5	
IR Core (Core to tank)	All TRF	>300	200 – 300 M Ohm	50 200	1 50	<1	2
HV Resistance between phases	All TRF	< 0.5%	0.5 - 1.0%	1.0 2.0%	2.0 - 3.0%	>3.0%	2
MV Resistance between phases	All TRF	< 1.5%	1.5-3.0%	3.0-4.0%	4.0-5.0%	>5.0%	2
LV Resistance between phases		< 1.5%	1.5-3.0%	3.0-4.0%	4.0-5.0%	>5.0%	
Bushing Tan Delta @ 20 Deg. C	All TRF	<0.5%	0.5 - 0.7%	0.7 - 1.5%	1.5 - 2.0%	>2.0%	2
Bushing Capacitance w.r.t. factory value / previous test results	All TRF	≤ ±5%	±5 - ±10%	±10 - ±15%	±15 - ±20%	> ±20%	2
Rate of rise of Bushing Tan Delta @ 20 Deg. C w.r.t previous results	All TRF	< 5%	5-10%	10-20%	20-35%	35-50%	2
SFRA		No deviation	Minor deviation	Moderate deviation	Significant deviation	Severe deviation	3
Transformer Service Factor (TSF)							
Load condition (in %)	All TRF	<50	50-80	80-100	100-120	>120	3
Age condition (in years)	All TRF	<20	20-25	26-35	36-40	>40	3
No of Faults fed by Transformer in the month (in Nos.)	All TRF	No fault	≤1	1-5	5-10	>10	1
Max Fault current seen by Transformer in the month (in %)	All TRF	≤25	25-50	50-90	90-95	> 95	1
Inspection & Maintenance							
Oil Leakage	All TRF	No leakage	Minor leakage which can be attended	Minor leakage which cannot be attended	Major leakage which can be attended	Major leakage which cannot be attended	2
Cooling fan / unit coolers condition	All TRF	All working and standby fans / unit coolers available	All working fans / unit coolers available	66 100% working / unit coolers fans available	33-66% working fans / unit coolers available	>33% working fans / unit coolers available	2
Reading of OTI (Maximum) in Deg.	All TRF	≤ 60	60-70	70-80	80-90	> 90	2
Reading of WTI (Maximum) in Deg.	All TRF	≤ 65	65-75	75-85	85-95	> 95	2
Thermo-vision scanning	All TRF	Normal	Minor defects found which can be attended	Minor defects found which cannot be attended	Major defects found which can be attended	Major defects found which cannot be attended	2

APPENDIX D. Some Aspects of Probability Theory

An independent variable is the one controlled in an experiment, e.g., the *height* above ground from which an object is dropped; the *time* it takes the object to hit the ground is a dependent variable.

The probability of an event, A, is the number of ways A can occur divided by the number of possible outcomes. The probability of rolling a 6 with a fair die is 1 in 6, or 16.67%. The sum of all outcomes must total 100%, or a probability of 1.0.

D.1 Independent Probabilities

Independent probabilities are where for two events, A and B, the occurrence of A does not affect the probability of B occurring. For example, the probability of rolling a 6 with a fair die is independent of the outcome of tossing a coin. In abbreviated form, if event A is rolling a 6 with a fair die, then the probability that event A occurs in a test (or trial, or experiment, etc.), P(A), is given by:

$P(A) = 1/6$ or $\sim 16.67\%$

If event B is flipping a coin and getting heads, then the probability of B occurring is P(B):

$P(B) = \frac{1}{2}$ or 50%

For independent events, the probability of both occurring is the product of their individual probabilities.

$P(A \text{ and } B) = P(A) . P(B)$

If the outcome of event A is known (i.e. the throw of the die), does that outcome affect outcome B? The answer is no, as they are independent. Therefore, the probability of B given the outcome of A, $P(B|A)$, is written just the same as the probability of B, which is P(B), which is 50%:

$P(B|A) = P(B) = 0.5$

Similarly, as A does not depend on B:

$P(A|B) = P(A) = 1/6$

D.2 Dependent or Conditional Probabilities

Conditional probabilities result when the outcome of one event influences the outcome of other events.

Simple examples of conditional probabilities often refer to drawing cards from a deck of cards or colored balls from a container[15]. For example, what are the chances of first drawing a queen from a

[15] Usually an urn, for some reason.

deck of cards, and then, without replacing the queen, drawing a jack? Let event A be the drawing of the first card as a queen and B be the drawing of the jack *given that a queen was drawn first.* Then:

P(A) = 4/52 = 1/13

P(B|A) = 4/51

(51 cards left as a queen was drawn for the first event; if a queen was not drawn then the chances are 0)

What is the probability that both events, A and B, will occur? It is the multiplication of the two separate probabilities to yield P (A and B):

P (A and B) = P(A) x P(B|A) = P(A). P(B|A) = 4/52 x 4/51 = 16/2652 = 4/663 = ~0.6%

Tree diagrams are useful to help understand conditional probabilities—they do not indicate the order of events, but list possible outcomes. For the two card drawing events, Queen and Jack, a tree diagram may be as follows:

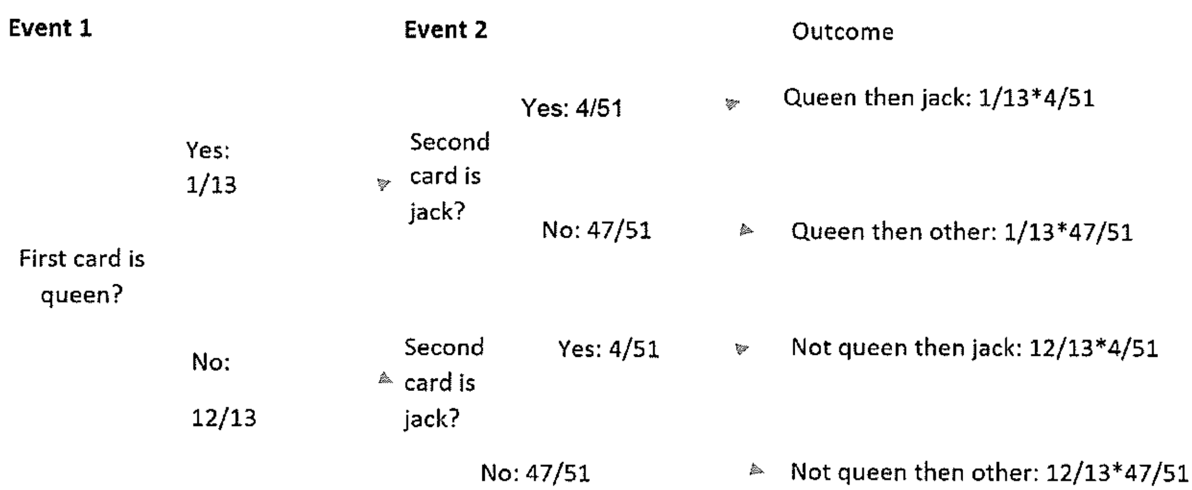


Figure D-1: Decision Tree for 2 Events

D.3 Bayes Theorem

The Rev. Bayes developed a theorem to calculate conditional probabilities when ______ (fill in the blank with a common subset of requirements). For more in-depth analysis, see [52]:

- The sample space may be partitioned into mutually-exclusive events (rains, does not rain).
- The probabilities evaluated for those exclusive events are available.
- There is an event B that is conditional on the exclusive events.
- There is the conditional probability for B for each mutually exclusive event.

In general terms for events A and B:

P(A | B). P(B) = P(B | A). P(A)

An example may help; a subsequent tree diagram may also clarify the example.

A particular test for drug abuse is 95% accurate—that is, if the test is administered correctly it will:

- Indicate drug use in a drug user 95% of the time (true positive)
- Indicate no drug use in a drug user 5% of the time (false negative)
- Indicate no drug use in a non-drug user 95% of the time (true negative)
- Indicate drug use in a non-drug user 5% of the time (false positive)

If a test indicates a positive result, and prevalence of drug use in the population is 1%, what is the probability that the person tested is, in fact, a drug user?

In a Bayesian analysis, look at the two mutually exclusive results, call them A1 and A2, and assign probabilities:

- A1: the person is a drug user, which occurs 1% of the time so P(A1) = 0.01
- A2: the person is not a drug user, which occurs 99% of the time so P(A2) = 0.99

Note that there is no A3, or A4 as A1 and A2 cover the complete possible range of drug user possibilities.

It is also known that the probability a test is correct is 95%; let B *be the event of a positive test* and the conditional probability of B given either A1 or A2 can be examined:

- P(B | A1) = 0.95 as for a drug user, will be accurate 95% of the time (true positive)
- P(B | A2) = 0.05 as for a non-drug user there is a 5% false positive rate

The probability of interest is P(A1 | B), which is the probability that a person actually does take drugs given that they test positive.

Use Bayes rule:

P(A1 | B). P(B) = P(B | A1). P(A1) or P(A1 | B) = P(B | A1). P(A1)/ P(B)

Now, find P(B) as that is not yet known. Use a common conditional expansion that covers the two mutually exclusive options A1 and A2:

P(B) = P(B | A1). P(A1) + P(B | A2). P(A2)

Which yields:

P(A1 | B) = P(B | A1). P(A1)/ [P(B | A1). P(A1) + P(B | A2). P(A2)]

All of the terms are in the right hand side of the equation; substituting yields:

P(A1 | B) = 0.95*0.01/ [(0.95*0.01) + (0.05 * 0.99)] = 0.16 or about 16%

This result may seem surprising. The prevalence of the drug use in the population significantly affects the seeming validity of the result.

A tree diagram, based on a population of 10,000 people, may help:

- Given the prevalence of drug use, 100 people are drug users, and 9,900 are not.
- For the 100 drug users, 95 will test positive and 5 will test negative.
- For the 9,900 non-drug users, 495 will test positive and 9,405 will test negative.
- Therefore, there are (95+495) = 590 people who will test positive.
- Of those 590 people, 95 are true positives.
- So, the chances of a drug user testing positive is 95/590 = 0.16 or about 16%.

In summary, as a decision tree:

Population:		Test		Outcome
10,000			Yes: 95	95 true positive
	Yes: 100	Test positive?		
			No: 5	5 false negative
Drug Use?				
	No: 9,900	Test positive?	Yes: 495	495 false positive
			No: 9,405	9,405 true negative

Figure D-2: Decision Tree for Drug Testing

This type of analysis is occasionally used to show the problem with drug testing, as there is only a 16% chance of a true positive being obtained when a positive result is found. The inverse is that a negative is overwhelmingly likely to be a true negative: 9405/9410 = 0.9995 or 99.95%.

Probability theory is often difficult to deal with in the abstract, but practical examples with real numbers and decision trees usually help clarify.

D.4 A Practical Example of Bad Math Probability Analysis

When Nico Rosberg retired from Formula 1 racing after winning the 2016 World Championship, there was much speculation as to his successor as Lewis Hamilton's teammate. The BBC website gave a summary[16] of possible drivers and their likely chances of getting the role. Only one driver can become the teammate.

Table D-1 shows a number of entries from the BBC article and their estimated probabilities of success.

Table D-1: Estimated Probabilities of Individual Driver Success

Driver	Probability of Success
Fernando Alonso	6/10
Daniel Ricciardo	6/10
Max Verstappen	4/10
Sebastian Vettel	8/10
Valtteri Bottas	9/10
Pascal Wehrlein	8/10

The problem in Table D-1 is that the sum of the individual and discrete outcomes—an individual driver's success—sums to far greater than 100%. Looking at each value individually, the probability estimates may seem reasonable, even justified, but the math is rigid, and the sum of values must be 100%. The values could be normalized or scaled to allow the summation to be correct, as shown in Table D-2.

Table D-2: Normalized Estimated Probabilities of Individual Driver Success

Driver	Probability of Success	Scaled Value	% Probability
Fernando Alonso	6/10	1.47/10	14.7
Daniel Ricciardo	6/10	1.47/10	14.7
Max Verstappen	4/10	0.96/10	9.6
Sebastian Vettel	8/10	1.95/10	19.5
Valtteri Bottas	9/10	2.20/10	22.0
Pascal Wehrlein	8/10	1.95/10	19.5

The scaled values in Table D-2 are not as spectacular, but they are, at least, realistic!

[16] www.bbc.com/sport/formula1/38185491

APPENDIX E. Combining Independent Probabilities

Independent factors are often assumed, but this may not be the case.

If the chance of rain on Saturday is 50% and the chance of rain on Sunday is 50%, then what are the chances it will rain this weekend (that is, on Saturday or Sunday or on both days)? A tree diagram helps identify the probabilities associated with outcomes.

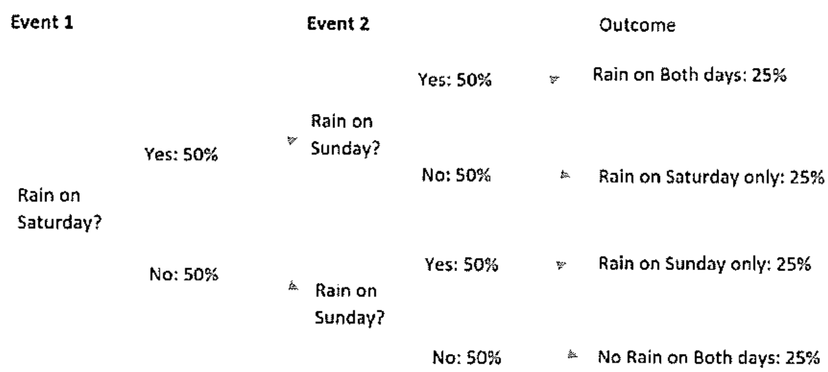


Figure E-1: Tree Diagram for Rain Probabilities

Event 1: "Rain on Saturday" has two outcomes, each assigned a probability of 50%.

Event 2: "Rain on Sunday" has two outcomes, each assigned a probability of 50%.

The tree indicates the 4 possible outcomes, with associated probability for that outcome.

The chance of rain on the weekend is 75%, obtained by summing the probabilities for the independent outcomes. Independence also means that the mathematician does not have to cater for the order of events.

The tree can be made more general by using:

- P1 for the probability of rain on Saturday, so (1-P1) for the probability of no rain on Saturday
- P2 for the probability of rain on Sunday, so (1-P2) for the probability of no rain on Sunday

The diagram now becomes as shown in Figure E-2:

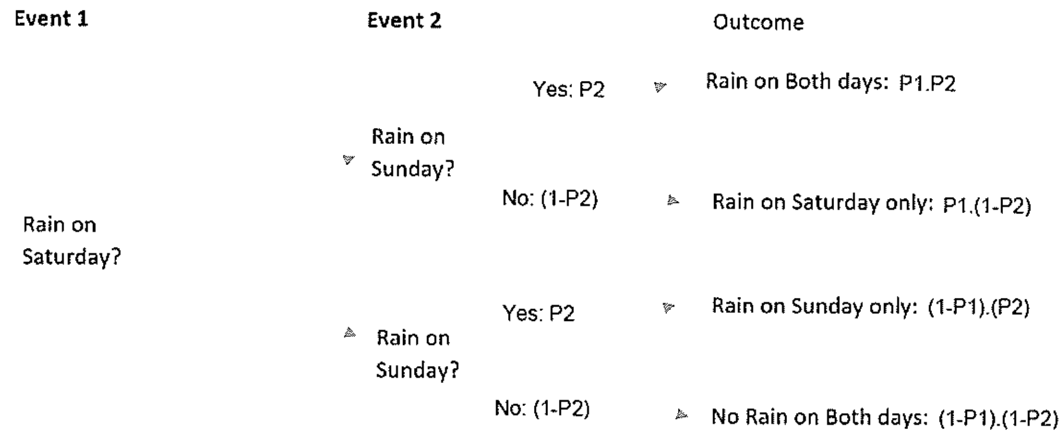


Figure E-2: Generalized Tree Diagram for Rain Probabilities

If the desired value is the probability of having rain at all, that is, on either or both days, then sum up all the probabilities for rain on different days, or use $(1\text{-}P1).\ (1\text{-}P2)$ to calculate the probability that there is no rain on any day, and subtract that from 1.

Thus:

$$\text{Probability of rain } P(R) = P1.\ P2 + P1.\ (1\text{-}P2) + (1\text{-}P1).\ P2$$

Or:

$$\text{Probability of rain } P(R) = 1 - ((1\text{-}P1).\ (1\text{-}P2))$$

This can be generalized to multiple days. If PN is the probability of rain on day N, then the probability of rain on at least one day is:

$$P(R) = 1 - ((1\text{-}P1).\ (1\text{-}P2) \ldots (1\text{-}PN))$$

How does this relate to PoF for an asset? If there are a number of independent factors, each of which has an estimate of probability of causing failure, then a probability that the asset will fail because of those factors can be calculated.

If the factors are identified by letter, A, B… and the probability of failure for each factor is designated P(A), P(B) etc., then use the same approach as for rain: $(1 - P(A))$ is the probability of surviving factor A.

The probability of surviving P(Survive) is given by:

$$P(\text{Survive}) = (1 - P(A)).\ (1 - P(B)).\ (1 - P(C)).\ (1 - P(D)) \ldots (1 - P(N))$$

And:

$$P(\text{Fail}) = 1 - P(\text{Survive})$$

This is the approach, used in Microsoft Excel™, to combine probabilities in Table E-1. For three independent factors,

Table E-1 shows some values for the individual probability of an event and the combined probability of at least one event occurring. If the event is failure, the P(Survive) and P(Fail) can be calculated.

Table E-1: Combining 3 Independent Probabilities

	P(A)	*P(B)*	*P(C)*	*Overall P(Survive)*	*Overall P(Fail)*
Probabilities	0.01	0.01	0.01	0.970	0.030
	0.02	0.02	0.02	0.941	0.059
	0.005	0.005	0.005	0.985	0.015
	0.9	0.9	0.9	0.001	0.999
	0.01	0.02	0.03	0.941	0.059
	0.1	0.2	0.3	0.504	0.496
	0.25	0.25	0.25	0.422	0.578
	0.002	0.002	0.1	0.896	0.104

Table E-1 shows that for three independent PoFs, the combined PoF has a value somewhere between the maximum of the individual factors and the sum of the factors.

Caveat: it is often assumed that factors are independent when they are, in fact, not. When looking at DGA values, there is a PoF which can be deduced based on hydrogen; the value for PoF for acetylene may also be calculated, but it is probably not independent of PoF based on hydrogen, as both may have a common cause and be a symptom of that cause. Much the same happens with rain—the likelihood of rain tomorrow correlates strongly with the actuality of rain today; the two are dependent.

APPENDIX F. On Time

This is a brief discussion of the role that time plays in asset condition and the development of an asset condition assessment. Asset condition is not a constant; thus, the estimate of the condition based on available data and analyses may change over time.

F.1 Effects of Time, and Resulting Relative Timescales

Time enters into consideration in a number of ways. Table F-1 indicates relative timescales which vary from short (milliseconds to seconds) to long (months to years). At this point, do not define actual timescales, as times for generating data, assessing data, and planning intervention may be short/medium/long-term varying. It is up to the relevant stakeholders to characterize in detail what these terms (short, medium, long, etc.) mean in their technical/business/asset management context.

Table F-1: Assets, Timescales and Variance

Case	Time Variance	Timescale for Variance
Asset condition	Asset condition is variable over time – usually deteriorating, not improving; the rate of change is a variable and may accelerate as we get closer to failure. Nowlan and Heaps [55] RCM curves may apply to an individual asset	Short – long:
Data	Some data will be relatively static – nameplate data, say; measured parameters may vary at different rates with time – PD values may vary extremely rapidly; the frequency with which measurements are made must be relevant to the dynamic nature of the data itself	Short – long: Depending on failure mode to be identified and tracked
Failure modes	Some deterioration will lead to failure modes evolving rapidly over time, accelerating, as it were, while other modes develop more gracefully; the rate of change of failure mode may imply a need for a change in generation or update of data	Short – long: Depending on evolution of failure mode
Data relevance	Measured values will have less relevance as they become older	Short – long: Depending on failure mode
Assessments	As conditions change and deterioration evolves, the need for more frequent assessments may follow; note an assessment is a more detailed review than a standard AHI	Updated on change of data which relates to failure mode and raw data
Planning Intervention	As condition deteriorates, we may need to respond in different timescales for different failure modes, depending on the suspected evolution of the failure mode and the increasing likelihood of failure or the inability for an asset to perform its function	Short – long: Depending on failure mode, rate of change of mode, and the asset management context
Asset Management Assessment review	The success of an assessment program itself should be reviewed over time to check that it is providing value	Medium – long: It is good practice to review performance of assessments in an appropriate timescale

F.2 Spectrum of Timescales

It may be considered that if there were no asset deterioration, there would be no condition-based failure modes to cause concern—failures would be governed by external causes of failure, such as may be related to weather, animal incursion, switching incidents, vandalism, etc. Further considerations, such as the role of design/manufacturer-specific issues in allowing specific failure modes within a transformer to develop in preference to others, may themselves be effectively constants; the result, however, is still deterioration.

Table F-1 considers the role of time in different cases and uses qualitative descriptions: short, medium, etc. Table F-2 looks at giving some ***possible*** values to those words across a spectrum of time.

It must be stressed that the actual timescales for action will depend on the failure modes considered, their timescales for evolution, and the business context of the individual organization.

Table F-2: Terms, Timescales, and Applications

Term	Timescale	Comment or application
Immediate	Within seconds	Protection would be included here; responding to high-level PD alarms or bushing monitoring
Very short	Within hours to days	Response to other high-level monitoring alert indicating rapidly-evolving and close to end-of-life condition
Short term	Days to 1 month	More gracefully evolving condition; thermal issues
Soon	Same as "Short term"	
Medium term	1 months to 1 year	Operational responses will have a different need than strategic responses; maintenance planning and capital programs overlap
Near long	1-5 years	
Long term	> 5 years	
Foreseeable future	>15 years	At this point, individual condition assessments have very little precision or accuracy and we're indulging in statistics

In Table F-2, the timescale terms are monotonically getting larger, but not uniformly—in linear, logarithmic, or exponential manners. For example, the Very Short timescale is effectively 1,000s of times larger than the previous category, Immediate. However, the Short Term may only be a few times larger than the Very Short. This lack of uniformity is not an issue, as long as the timescales are understood and appropriate actions and responses are planned and carried out as necessary.

Each organization must come up with generic and calibrated timescales such that they have both clear and consistent meaning in their application across the organization.

Failure to do so will lead to an inconsistency of results and inability to justify decisions under review or audit.